Hybrid Deep Learning Approach for Monthly Rainfall Prediction Using Endogenous property

Presented by

Mandala Venkata Surendra
17CE33003

Under Supervision of
prof . Rajib Maity

Department of Civil Engineering
IIT Kharagpur

Introduction

- Rainfall prediction
  - ➢ Rainfall prediction is clearly of great importance for India
  - ➢ The spatio-temporal distribution of precipitation is getting modified as an impact of changing climate
- One would like to make
  - ➢ long term prediction, i.e. predict rainfall a few weeks or months in advance
  - ➢ short term prediction, i.e. predict rainfall over different locations a few days in advance

Problem Statement

- Among various spatio-temporal scales, seasonal or monthly prediction of rainfall over a subdivision prediction is one of the most important tasks
- How to capture its endogenous properties of hydrological time series with respect to their temporal evolution?
- Is the variation of rainfall solved using advanced algorithms better than the existing prediction methods?

Objective

- **Specific objective**
  - ➢ Capture the endogenous properties of sub-divisional monthly rainfall series.
  - ➢ Extract the hidden sequential information in the sub-divisional monthly rainfall series.
  - ➢ Develop monthly rainfall prediction model with a goal for improved prediction performance.

- **General objective**
  - ➢ Implementing of rainfall prediction system using hybrid DL, a combination of one-dimensional Convolutional Neural Network (Conv1D) and Multi-Layer Perceptron (MLP) for monthly rainfall prediction of different subdivisions of India.

Methodology

- The methodologies used in this study are :
  - ➢ Literature survey
  - ➢ Propose the system
  - ➢ Design and Implement
  - ➢ Testing

Contribution of this study

- To enhance the prediction mechanisms in different subdivisions of India by considering an appropriate model.
- To help for selecting and implementing of appropriate prediction algorithms for spatio-temporal distribution of precipitation in India.

Literature Review

- Artificial Neural Network based forecasting of consecutive rainfalls.
  - ➢ In 2018, Lee et al.:“Application of artificial neural networks to rainfall forecasting in the geum river basin, korea”
- DL based deep network algorithm for forecasting next day precipitation using environmental factors.
  - ➢ In June 2017,Zhang et al.: “A deep-learning based precipitation forecasting approach using multiple environmental factors”

Data Description

- The dataset consists of the monthly rainfall for the period 1901-2015 for each state in India
- The dataset contains 19 attributes (individual months, annual, and combinations of three consecutive months)
- The data from 1951–2000 were used to train the models, and the data from 2001–2015 were used for testing
- Previous months rainfall data is used as input to predict the consecutive month rainfall

Data description......

- Three different lags, i.e., 3, 6 and 9 months are considered
- Lag indicates the gap (number of months) between input i.e. past data and the starting month of prediction.

Data preprocessing

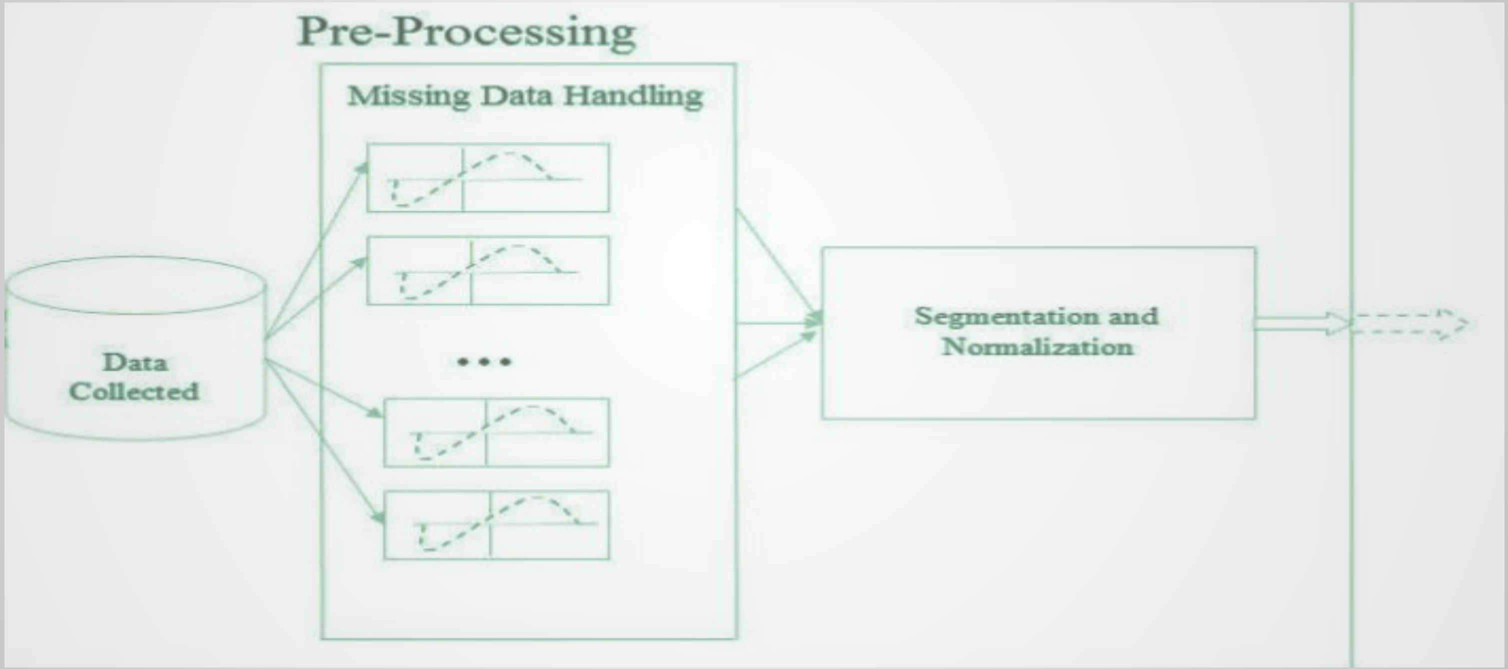


Figure 1: Block diagram of pre processing phase

preprocessing….

❑ Correlation among the variables

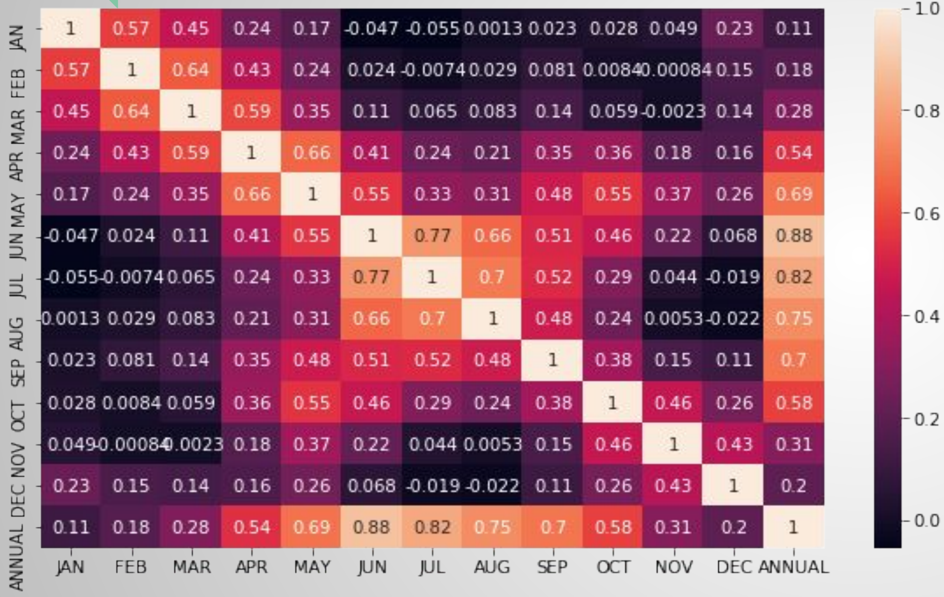


Figure 2. Heat map showing the correlation coefficients between rainfall in different months and the same with annual rainfall

Figure 3. Heat map showing the correlation coefficients between rainfall in different seasons and the same with annual rainfall

Model Architecture

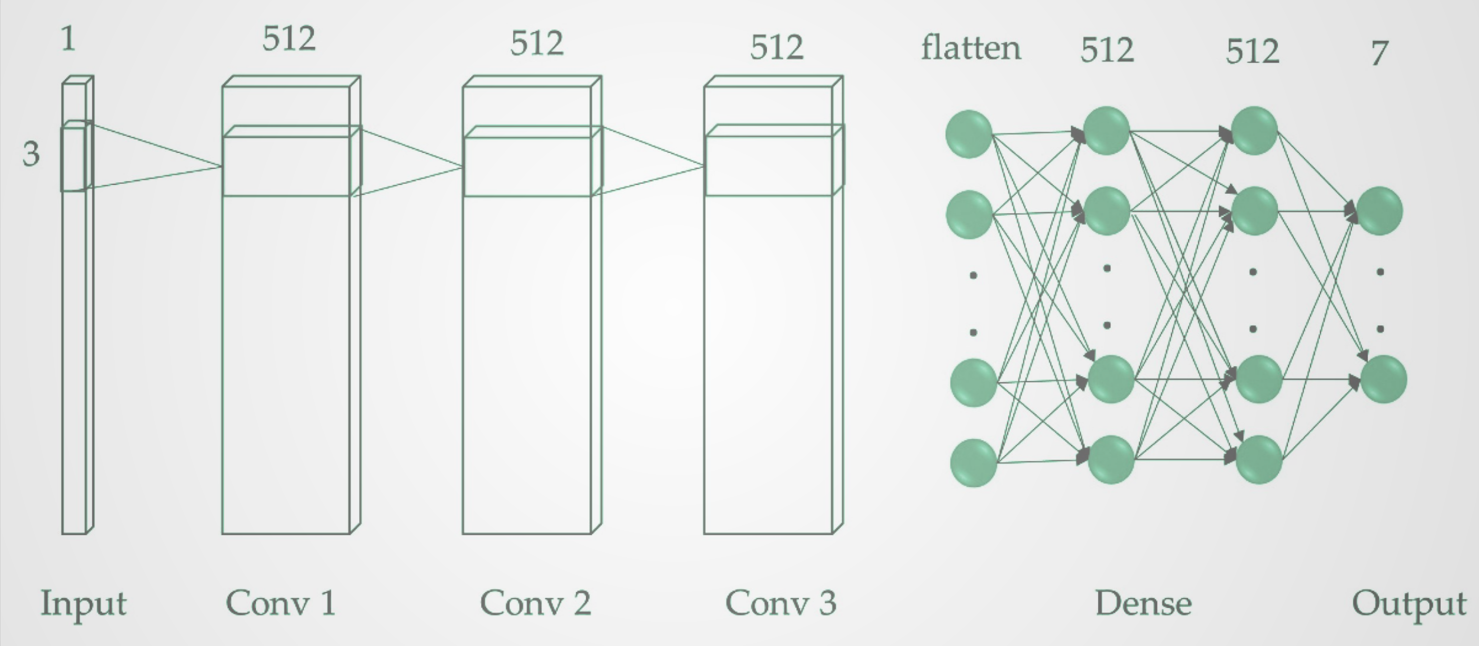


Figure 4: A sample 1D CNN configuration with 3 CNN and 2 MLP layers

Algorithm : CNN Training Procedure

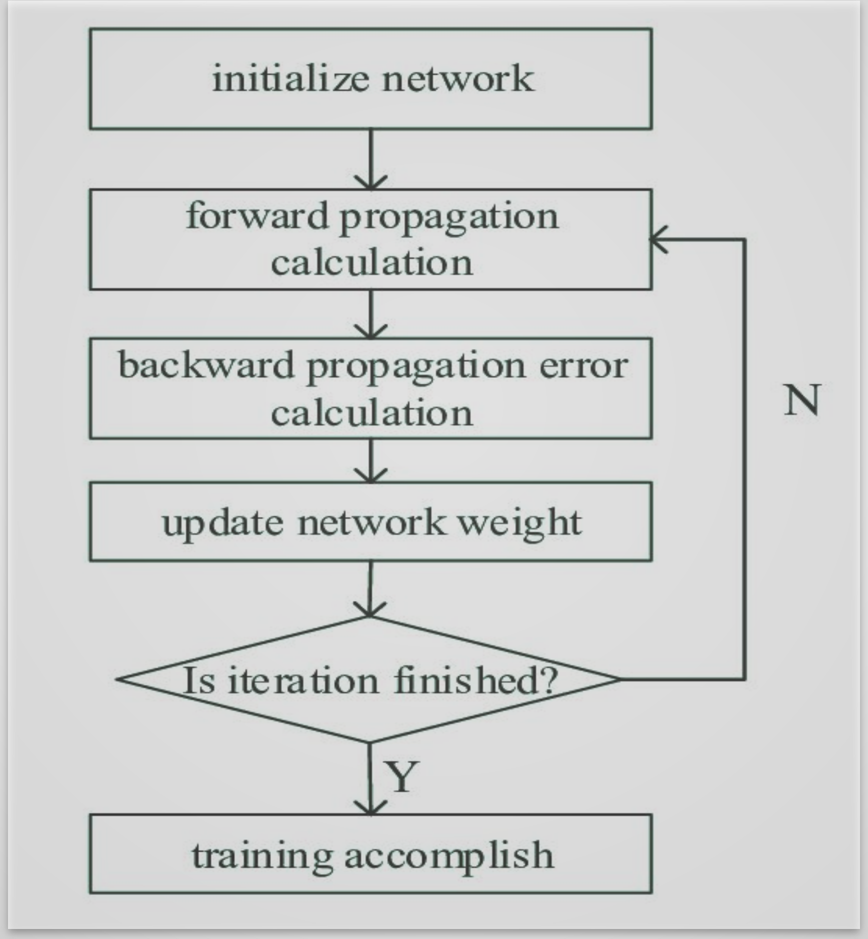
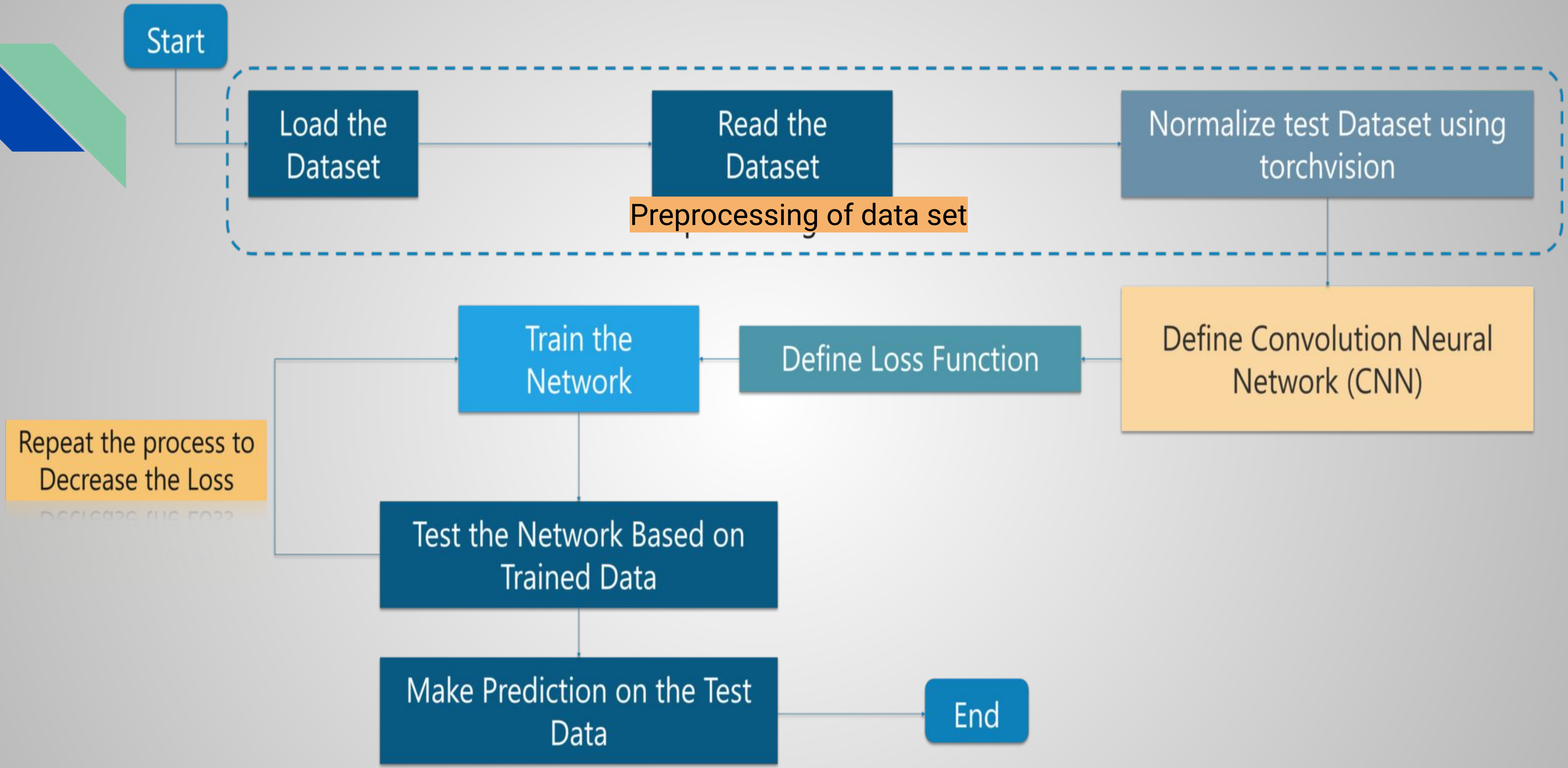


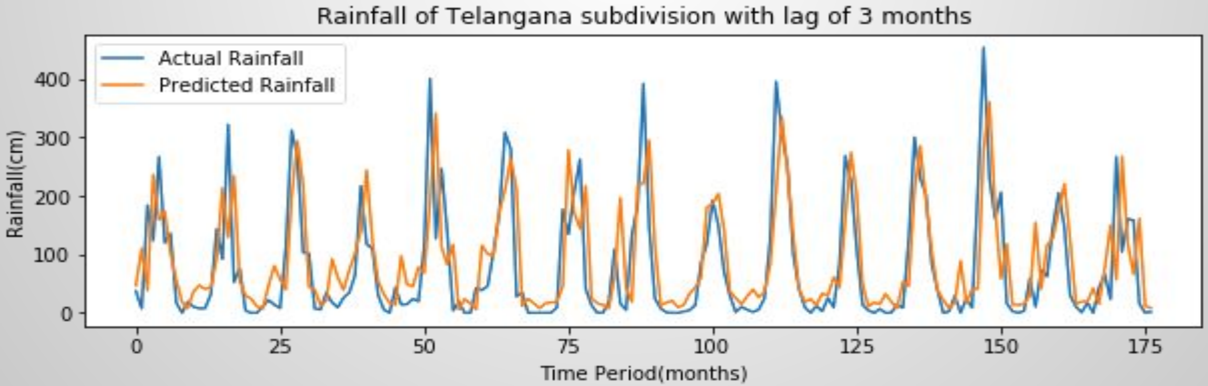
Table 1: Configurations of proposed hybrid Conv1D-MLP model

Layer no.	Layer	Type	Parameters of layers			
			Activation func.	Kernel Size	No.of filters	Neurons
1	Conv1D	Convolution Layer	ReLU	1	64	-
2	Conv1D	Convolution Layer	ReLU	2	128	-
3	Flatten	Flatten Layer	-	-	-	-
4	dense	Fully connected Layer	ReLU	-	-	128
5	dense	Fully connected Layer	ReLU	-	-	64
6	dense	Fully connected Layer	ReLU	-	-	32
7	dense	Fully connected Layer (output layer)	Linear	-	-	1


Start
Load the Dataset
Read the Dataset
Normalize test Dataset using torchvision
Preprocessing of data set
Define Convolution Neural Network (CNN)
Define Loss Function
Train the Network
Repeat the process to Decrease the Loss
Test the Network Based on Trained Data
Make Prediction on the Test Data
End

Results and Discussion

- Trained model with the aforementioned architecture is tested on three subdivisions of India, observed the predicted rainfall for all the months from 2001-2015
- **3 months Lag**



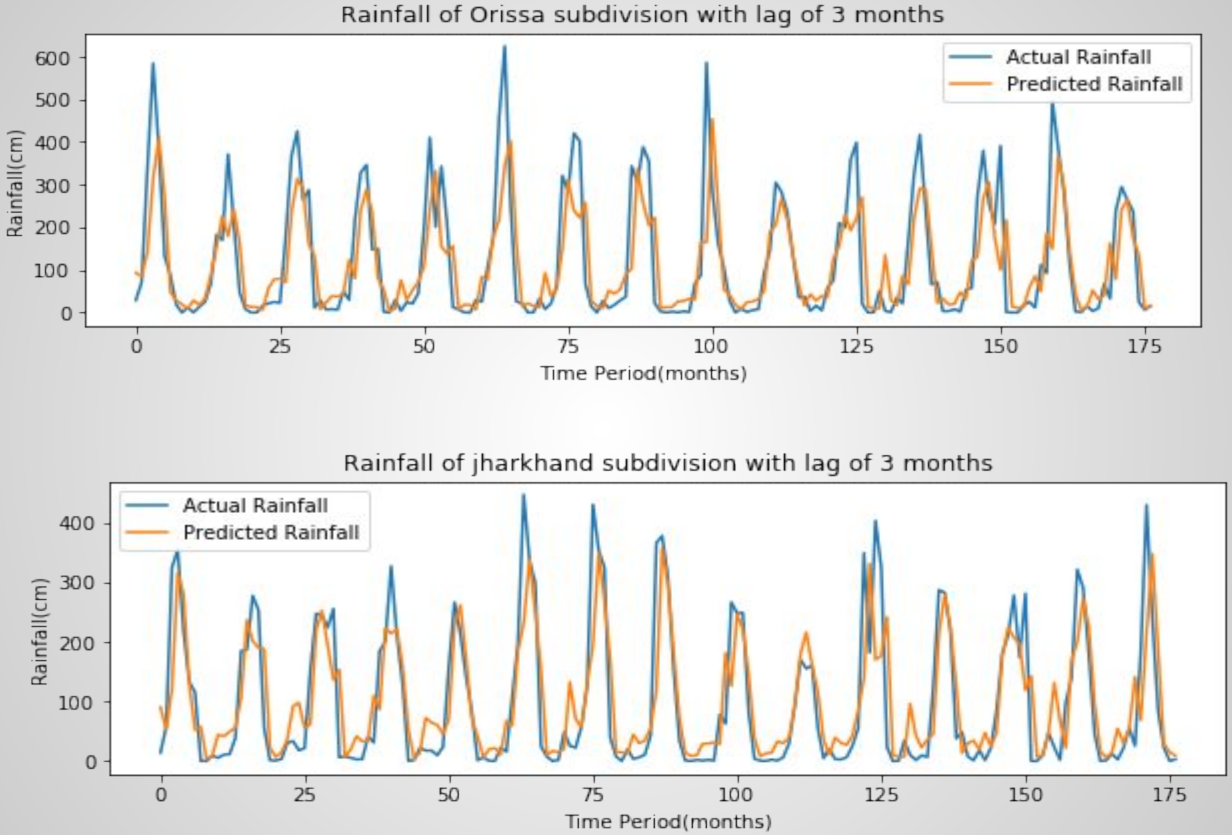
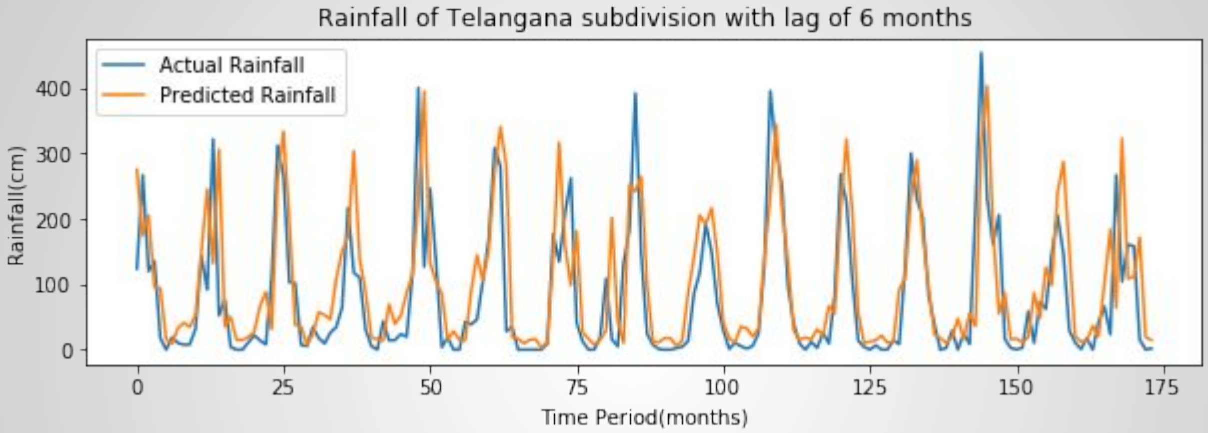
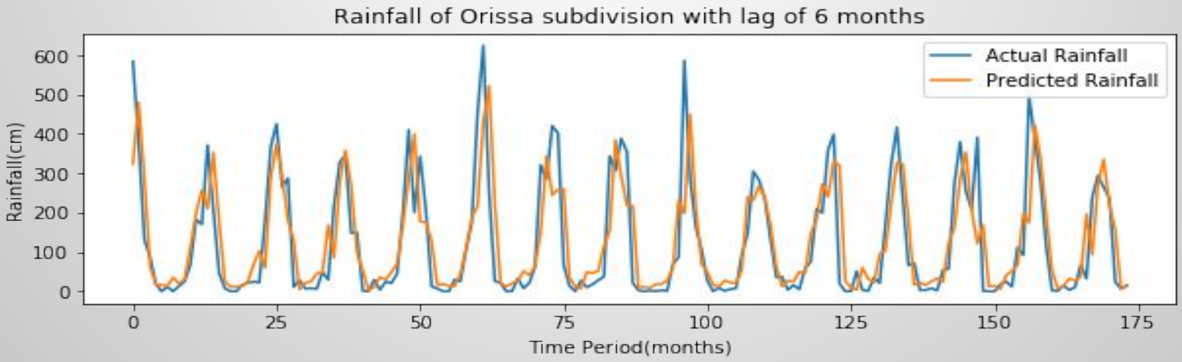


Figure 5. Comparison plots between actual and predicted monthly using inputs from 3 previous months for 3 different subdivisions. X-axis shows the months from the year 2001 to 2015.

❑ 6 months lag



Rainfall of Telangana subdivision with lag of 6 months
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
0
100
200
300
400
0
25
50
75
100
125
150
175
Time Period(months)



Rainfall of Orissa subdivision with lag of 6 months
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
0
100
200
300
400
500
600
0
25
50
75
100
125
150
175
Time Period(months)

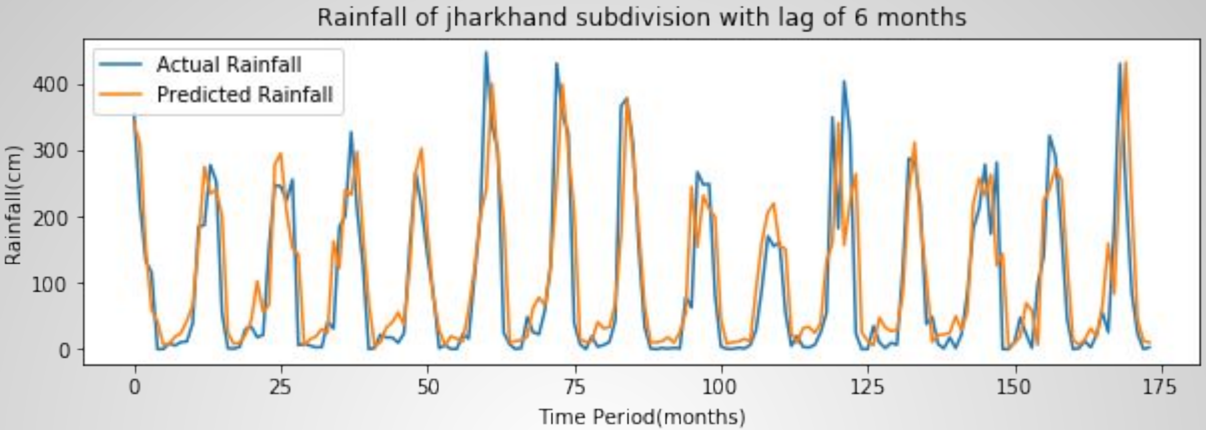
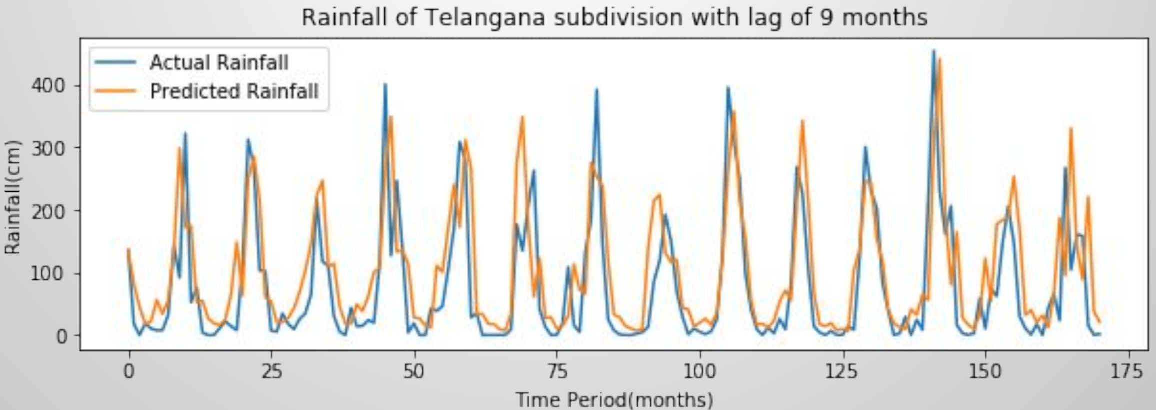


Figure 6. Comparison plots between actual and predicted monthly using inputs from 6 previous months for 3 different subdivisions. X-axis shows the months from the year 2001 to 2015.

- **9 months lag**



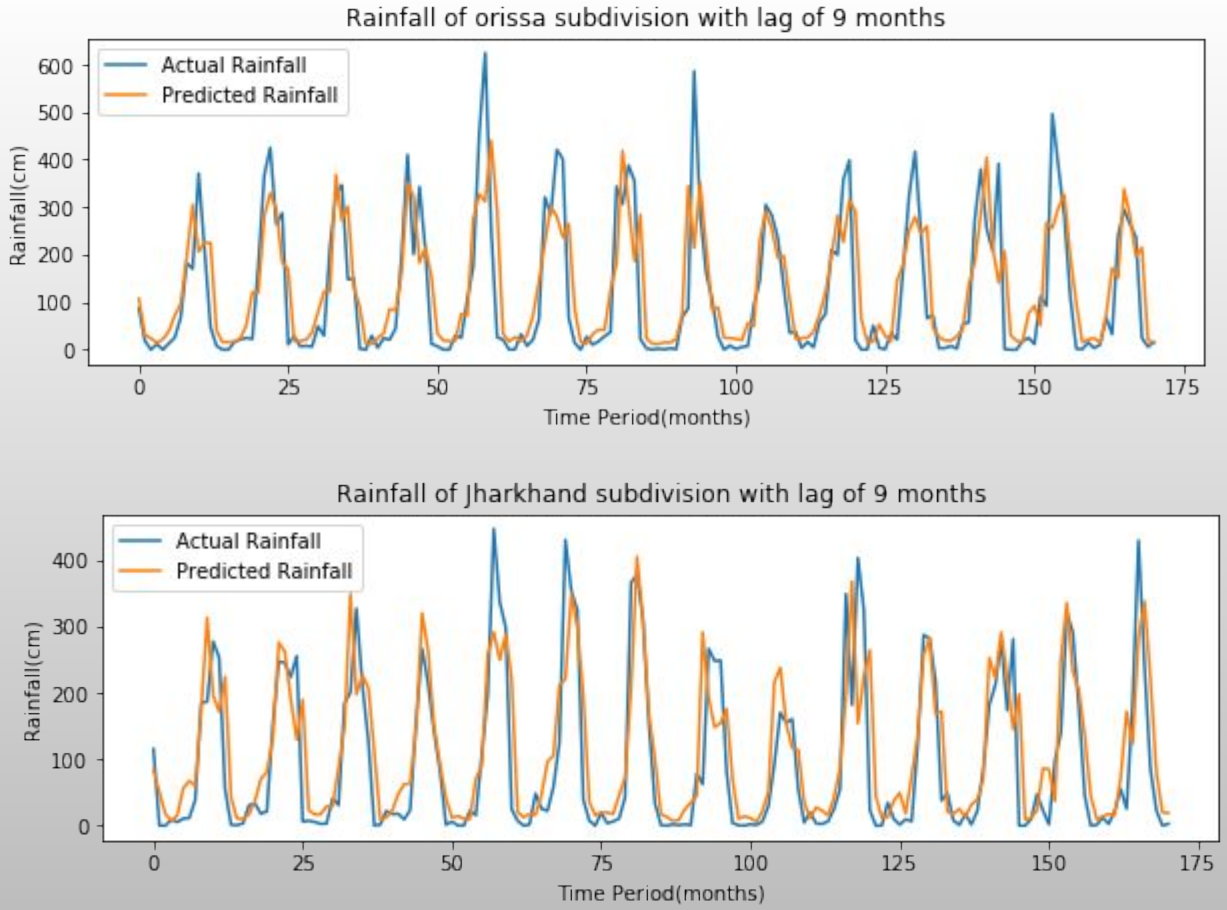


Figure 7 Comparison plots between actual and predicted monthly using inputs from 9 previous months for 3 different subdivisions. X-axis shows the months from the year 2001 to 2015.

TABLE 2: Performance statistics viz. r, RMSE and NSE for different lag periods at different subdivisions.

Subdivision	No.of inputs/ Lag months	Performance Statistics		
		RMSE	Correlation coefficient(r)	NSE coefficient
Telangana	3	74.25	0.686	0.445
	6	74.10	0.708	0.399
	9	73.36	0.728	0.459
Orissa	3	100.74	0.747	0.545
	6	95.68	0.770	0.590
	9	92.2	0.773	0.597
Jharkhand	3	71.87	0.803	0.642
	6	68.8	0.824	0.670
	9	67.65	0.806	0.676

- Increase in the performance with the increase in no. of inputs.
- Coefficients of correlation considering all the tested subdivisions lie between 0.6 - 0.8, NSE coefficient is greater than which indicates the model with different input sets gives good performance
- But with 9-month lag input best results are obtained
- Model yields better results for some of the subdivisions that is reflected through the above performance statistics table

Conclusion

- Hybrid Conv1D-MLP model has potential for monthly rainfall prediction of daily rainfall using past data as the input variables
- From performance it is concluded that deep learning has the potential to capture the non-linear relationship between the past data and monthly rainfall variability
- So the predictions obtained from the proposed hybrid DL model can be helpful in agriculture, flooding due to heavy rainfall etc..

Future Work

- Application to other Subdivisions
- Month-wise analysis of predictability
- Comparison with other machine learning models, such as support vector regression(SVR)
- Spatial variation in long-lead predictability of monthly rainfall using global climatic indices

THANK YOU!