Rainfall prediction for the state of Gujarat using deep learning technique

Rushikesh Nalla¹ and Urmil Kadakia¹ Supervisor: Prof. Ranendu Ghosh

Email: \$\frac{1}{2}01401106@daiict.ac.in}\$\$
201401013@daiict.ac.in

DAIICT, Gandhinagar- 382007, Gujarat, India

B.Tech Term Project, 2018



Outline

- Introduction
- 2 Literature Review
- Problem Statement
- 4 Features
- Rainfall sources
- 6 Problems in AWS Rainfall Data
- Datasets
- B Dataset Processing
- Implementation
- Multi layer Perceptron (MLP)
- Long short-term memory (LSTM)
- Results
- Conclusion
- Future Work

Introduction

- Predicting rainfall is extremely challenging.
- Rainfall varies both spatially and temporally.
- Precise amount of rainfall depends on a number of factors.
- Different ways of predicting rainfall:
 - Satellite Images
 - Digital Images
 - Weather radar maps

Literature Review

- Using only Infrared (TIR1) based algorithms, it incorrectly classify cirrus clouds as rain barring cloud.
- Passive Microwave is better compared to Infrared but it has low sampling frequency, thus low temporal resolution.
- CMORPH and MIRA uses both Infrared and Passive Microwave but has low temporal resolution due to PMW.

Literature Review (contd.)

- Visual and Infrared channels have high temporal resolution but visible data is not available at night.
- GMSRA uses information from 5 satellite channels coupled with pre-calibrated probability of rain derived from clouds top brightness temperature groups.
- PERSIANN-MSA uses self-organizing map(ANN-SOFM).

Problem Statement

Limitations of previous methods -

- Use of only one or two spectral channels.
- Daily or monthly average rainfall prediction.
- Very few studies focused on Indian regions.

Our Approach -

We have used multi-spectral channels with hourly predictions over $10km \times 10km$ for the state of Gujarat having coordinates 68W-75E and 25N-20S.

Features

- INSAT-3D is a geosynchronous satellite with IMAGER and SOUNDER payloads.
- VHRR data of multiple wavelengths has been taken from MOSDAC for each hour of each day for rainy months i.e. June-September from 2014 to 2017.
- 5 channels out of the 25 available spectral channel are used because other channels are not related to the rain or not available.

Features (contd.)

- Wavelength range of each channel:
 - \triangleright 0.52 μ m-0.72 μ m VIS (Visible)
 - ▶ 1.55 μ m-1.70 μ m SWIR (Short Wave Infrared)
 - ▶ 6.50μ m- 7.00μ m WV (Water Vapor)
 - ▶ 10.2μ m- 11.2μ m TIR-1 (Thermal Infrared)
 - ▶ 11.5μ m- 12.5μ m TIR-2 (Thermal Infrared)
- Temporal resolution: 1 hour
- Spatial resolution: 2×2 km

Features (contd.)

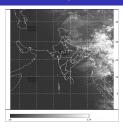


Figure 1: VIS $(0.65 \mu m)$



Figure 2: SWIR $(1.625 \mu m)$

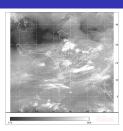


Figure 3: WV (6.8 μ m)

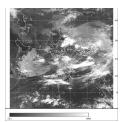


Figure 4: TIR1 ($10.8\mu m$)

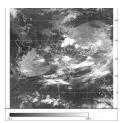


Figure 5: TIR2 ($12\mu m$)

Features (contd.)



Figure 6: SRTM DEM



Figure 7: NDVI

- SRTM provides Digital Elevation data with 1km×1km resolution and the vertical error less than 16m.
- \bullet The NDVI is a numerical indicator that shows whether targeted area contains vegetation or not. it has 15 days temporal and $1 km \times 1 km$ spatial resolution.

Rainfall sources

- Ground based rainfall observation stations (AWS data) provided by MOSDAC, ISRO.
- IMD rainfall data (daily average).
- ullet TRMM PR hourly data with spacial resolution of $10 km \times 10 km$.

Problems in AWS Rainfall Data

- Rainfall data is cumulative and the reset points are at random.
- Appearance of 1023 due to two reasons:
 - ▶ It is the highest value (10 bit number).
 - ▶ Missing data is sometimes filled with value 1023.
- Random length of increasing and decreasing numbers.
- Garbage values like 9999 appear randomly.
- Nearly $(1/3)^{rd}$ data is missing.

Datasets

Da	ataset	Channel	Features per channel			Resolution	
	1	TIR1 (T)	5	SRTM, NDVI, WV (rad)	8	2×2 km	
	2	VIS (Albedo), SWIR (Rad), WV (T), WV (Rad), TIR1 (T), TIR2 (T)	5	SRTM, NDVI	32	10×10 km	

Table 1: Features in each dataset

Dataset Processing

- Computing the brightness temperature mean and standard deviation of 3×3 neighbourhood for each channel.
- Computing the brightness temperature mean and standard deviation of 5×5 neighbourhood for each channel.
- SWIR and VIS are highly affected by solar zenith angle (SZA) so we adopted a correction technique of multiplying each observed value with $(cosSZA)^{-1}$

Implementation

- Started with some linear classifier models but due to dis-satisfactory results we moved on to non-linear classifiers.
- 80% of total data was allocated for training and rest (20%) was used for validation.

class ID	Rainfall(in mm)
class 0	0
class 1	<2
class 2	<5
class 3	>=5

Table 2: Class ID and its corresponding rainfall (in mm) range

Multi layer Perceptron (MLP)

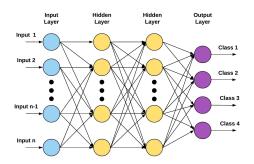


Figure 8: Multi Layer Perceptron

Input units = Number of features Output units = 4 (one for each class) Optimizer = Adam Optimizer Neurons in each hidden layer = 16*(number of input units)
Loss Function = Categorical cross-entropy
Activation Function = relu (hidden layers) and softmax (output layer)

Multi layer Perceptron (MLP) (contd.)

Synthetic Minority Oversampling Technique (SMOTE):

- SMOTE is a synthetic data generation technique which helps in balancing classes.
- If the classes are not balanced then the classifier always learns to predict the majority class.
- To avoid this and do the rightful prediction of rainfall we use SMOTE.

Long short-term memory (LSTM)

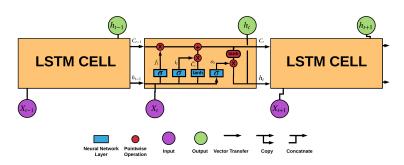


Figure 9: Long Short Term Memory Module

Input units = Number of features. For dataset 1 first Layer = 24 LSTM cells Next layer = time distributed dense layer Loss Function = Categorical cross-entropy Output layer = 4 (one for each class)
For dataset 2 First Layer = 10 LSTM
Optimizer = Adam Optimizer
Activation Function = softmax for output layer

Metrics

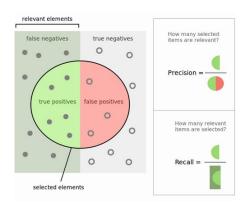


Figure 10: Precision and Recall

$$Accuracy = \frac{true \; positive \; + \; true \; negatives}{true \; positive \; + \; false \; negative \; + \; false \; positive \; + \; true \; negatives} \qquad \textit{Fscore} = \frac{2 \times \textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

Results

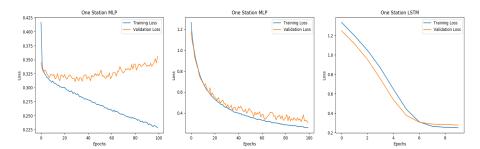


Figure 11: MLP model for SAC BOPAL without SMOTE SAC BOPAL with SMOTE

Figure 13: LSTM model for SAC BOPAL

Results (contd.)

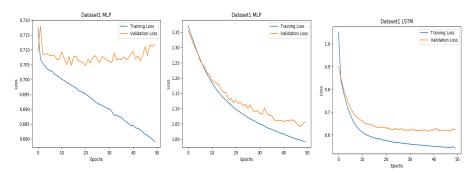


Figure 14: MLP model for Gujarat without SMOTE using Dataset 1

Figure 15: MLP model for Gujarat with SMOTE using Dataset 1

Figure 16: LSTM model for Gujarat using Dataset 1

Results (contd.)

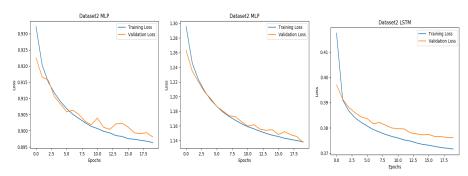


Figure 17: MLP model for Gujarat without SMOTE using Dataset 2

Figure 18: MLP model for Gujarat with SMOTE using Dataset 2

Figure 19: LSTM model for Gujarat using Dataset 2

Results (contd.)

(Da	ataset 1		Dataset 2					
	MLP wihout		LSTM	MLP without	MLP with	LSTM	MLP without	MLP with	LSTM
	SMOTE	SMOTE	LSTW	SMOTE	SMOTE	LSTIVI	SMOTE	SMOTE	LSTIVI
categorical_accuracy	0.93	0.91	0.95	0.80	0.5	0.85	0.62	0.46	0.84
val_categorical_accuracy	0.92	0.90	0.95	0.80	0.54	0.83	0.63	0.47	0.84
Precision overall	0.93	0.87	-	0.80	0.66	-	0.64	0.65	-
Recall overall	0.93	0.83	-	0.80	0.37	-	0.60	0.23	-
F-score overall	0.93	0.85	-	0.80	0.47	-	0.62	0.34	-
Precision_class1	0.00	0.79	-	0.50	0.62	-	0.66	0.68	-
Recall_class1	0.00	0.45	-	0.00	0.27	-	0.12	0.16	-
F-score_class1	0.00	0.57	-	0.00	0.38	-	0.20	0.26	-
Precision_class2	0.00	0.79	-	0.00	0.66	-	0.65	0.69	-
Recall_class2	0.00	0.67	-	0.00	0.37	-	0.01	0.16	-
F-score_class2	0.00	0.72	-	0.00	0.47	-	0.02	0.26	-
Precision_class3	0.00	0.83	-	1.00	0.77	-	0.68	0.78	-
Recall_class3	0.00	0.71	-	0.00	0.54	-	0.01	0.25	-
F-score_class3	0.00	0.77	-	0.00	0.63	-	0.01	0.38	-

Table 3: Results

Conclusion

- This project discusses the importance of different parameters that determine the amount of rainfall occurring in a particular region.
- Using multispectral channels (in dataset 2) over TIR1 channel (in dataset 1) didn't improve the results significantly as stated in some papers.
- MLP gave good results with SMOTE technique for a particular station data (small dataset).
- LSTM performed better than MLP in general as it made predictions with an accuracy of 84% when relatively balanced dataset was used.

Future Work

- Currently we used only 2014 year data for preparing dataset 2 and training the models due to limitation of computing resources.
- Here have only focused on Gujarat, like to develop this model for entire India.
- The IR satellite data taken from MOSDAC not entirely accurate so would try to incorporate data from NASA as well.
- Addition of Passive Microwave can be useful as it is directly related to the water content in the clouds.

Acknowledgement

We thank Mr. Arjun Bhasin, Head of Research and Development, Amnex InfoTechnologies Private Limited, for providing us with the opportunity to work on this project. We appreciate his constant support and guidance. We also thank Shivani Shah, Senior Scientist at ISRO, SAC for her technical assistance in acquisition and understanding of INSAT-3D data.

Reference

- Tim Bellerby, Martin Todd, Dom Kniveton, and Chris Kidd. "Rainfall estimation from a combination of TRMM precipitation radar and GOES multispectral satellite imagery through the use of an artificial neural network". In: Journal of applied Meteorology 39.12 (2000). pp. 2115–2128.
- MOSDAC INSAT-3D, Indian Space Research Organization(ISRO). https://www.mosdac.gov.in.
- Normalized Difference Vegetation Index (NDVI), SDAPSA, National Remote Sensing Center, Bhuvan Noeda. http://bhuvan.nrsc.gov.in.

 - Precision and Recall. https://en.wikipedia.org/wiki/Precision_and_recall.

Reference (contd.)

- Ali Behrangi, Kuo-lin Hsu, Bisher Imam, Soroosh Sorooshian, and Robert J Kuligowski. "Evaluating the utility of multispectral information in delineating the areal extent of precipitation". In: *Journal of Hydrometeorology* 10.3 (2009), pp. 684–700.
- Andy Jarvis, Hannes Isaak Reuter, Andrew Nelson, Edward Guevara, et al. "Hole-filled SRTM for the globe Version 4". In: available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org) 15 (2008).
 - Jesse Davis and Mark Goadrich. "The Relationship Between Precision-Recall and ROC Curves". In: Proceedings of the 23rd International Conference on Machine Learning. ICML '06. ACM, 2006, pp. 233–240. ISBN: 1-59593-383-2. DOI: 10.1145/1143844.1143874. URL: http://doi.acm.org/10.1145/1143844.1143874.

Dataset2 Distribution

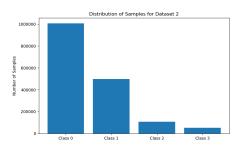


Figure 20: Sample Distribution for Dataset 2

AWS Data Problems

- Rainfall data is cumulative and the reset points are at random.
- Random length of increasing and decreasing numbers.
- Lot of missing data

@STATION_ID	LATITUDE	LONGITUDE	ALTITUDE(m)	TIME(GMT)	DATE(GMT)	AIR_TEMP(°C)	WIND_SPEED(m/s)	WIND_DIRECTION(deg)	ATMO_PRESSURE(hpa)	HUMIDITY(%)	RAIN_FALL(mm)
ISRO0009_15F009(PRL. Ahmedabad)	23.0362	72.5424	7	0	08/08/2016	24.9	0.1	264.9	1000.3	100	208
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	1	08/08/2016	24.7	(359.2	1000.6	100	208
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	2	08/08/2016	24.9	0.3	210.2	1001.2	100	209
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	5	08/08/2016	26	0.8	225.8	1001.9	100	214
ISRO0009_15F009(PRL. Ahmedabad)	23.0362	72.5424	. 7	5	08/09/2016	27.1	1.4	243.9	1001	100	222
ISRO0009_15F009(PRL. Ahmedabad)	23.0362	72.5424	. 7	6	08/09/2016	27.1	2.5	224.8	1000.8	100	222
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	7	08/09/2016	27.1	0.6	213.1	1000.4	100	222
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	17	08/09/2016	26	0.2	200.9	1000.2	100	11
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	20	08/09/2016	25.9	0.7	252.2	999.5	100	11
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	3	08/10/2016	25.1	12	230.2	1000.3	100	34
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	4	08/10/2016	25.5	1	213.1	1001	100	37
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	20	08/10/2016	25.5	(359.2	1002.1	100	43
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	21	08/10/2016	25.4	0.4	232.2	1001.2	100	43
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	22	08/10/2016	25.4	1.5	263	1000.8	100	43
ISRO0009_15F009(PRL. Ahmedabad)	23.0362	72.5424	7	23	08/10/2016	25.3	(359.2	1001.3	100	43
ISRO0009_15F009(PRLAhmedabad)	23.0362	72.5424	7	0	08/11/2016	25.2	0.2	159.8	1001.6	100	43

Figure 21

AWS Data Problems (contd.)

- Appearance of 1023 due to two reasons:
 - It is the highest value (10 bit number).
 - ▶ Missing data is sometimes filled with value 1023.

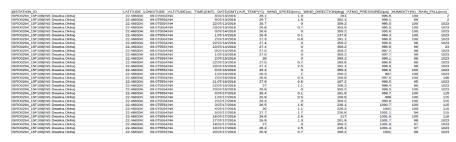


Figure 22

AWS Data Problems (contd.)

• Garbage values like 9999 appear randomly.

#STATION_ID			ALTITUDE(m)				ND_DIRECTION(deg)	ATMO_PRESSURE(hpa)		
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	20 07/10/2016	28.4	9955	0	9999.9		9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	21 07/10/2016	28.3	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	22 07/10/2016	28.8	1.4	276.1			6
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	23 07/10/2016	28.7	1.3	253.2			60
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	0 07/11/2016	28.4	1	247.8			68
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	1 07/11/2016	28.3	0.8	236.1	997.4	82	68
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	2 07/11/2016	28.6	1	236.1	997.6	81	68
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	3 07/11/2016	26.8	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	4 07/11/2016	30.3	1.2	237			68
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	5 07/11/2016	28.3	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	6 07/11/2016	28.2	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	7 07/11/2016	28.7	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	8 07/11/2016	29.2	9955	0			9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	9 07/11/2016	29.6	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	10 07/11/2016	29.3	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	11 07/11/2016	29.4	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	12 07/11/2016	29.4	9955	0			9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	13 07/11/2016	28.8	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	14 07/11/2016	28.8	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	15 07/11/2016	29.2	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	16 07/11/2016	29.3	9955	0	9999.9	35.1	9999.1
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	17 07/11/2016	29.3	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	18 07/11/2016	29.2	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	19 07/11/2016	29.1	9955	0	9999.9	35.1	9999.
SRO1024 15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	20 07/11/2016	28.6	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	21 07/11/2016	28.4	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	22 07/11/2016	28.3	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	23 07/11/2016	28.3	9955	0	9999.9	35.1	9999.
SRO1024_15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	0 07/12/2016	26.7	0.3	215.1	997.8	99	8
SRO1024 15F400(Jagudan (Main Spices Research Station))	23.513332	72.39861	87	1 07/12/2016	0	0	0	9999.9	35.1	9999.

Figure 23

AWS Data Problems (contd.)

- 95% of the entries have zero rainfall.
- AWS and IMD Rainfall don't even match.

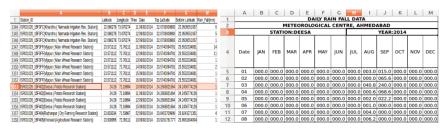


Figure 24

Abbreviation

- VIS = Visual Wavelength (1.55 μ m 1.70 μ m)
- MIR = Mid wave Infrared (0.39 μ m 0.7 μ m)
- WV = Water Vapor Wavelength (6.5 μ m 7.0 μ m)
- TIR1 = Thermal Infrared (10.2 μ m 11.2 μ m)
- TIR2 = Thermal Infrared (11.5 μ m 12.5 μ m)
- PMW = Passive Microwave (1cm 1m)
- VHRR = Very High Resolution Radiometer
- SRTM = Shuttle Radar Topography Mission
- DEM = Digital Elevation Models
- NDVI = Normalized Difference Vegetation Index
- AWIFS = Advanced Wide Field Sensor
- MLP = Multi Layer Perceptron

Abbreviation (contd.)

- LSTM = Long-Short Term Memory
- TRMM = Tropical Rainfall Measuring Mission
- AWS = Automatic Weather Station
- SZA = Solar Zenith Angle
- CMORPH = Climate Prediction Center morphing method
- MIRA = microwave/infrared rainfall algorithm
- GMSRA = Geostationary Operational Environmental Satellite system-Multispectral Rainfall Algorithm
- PERSSIAN-MSA = Rainfall prediction from Remotely Sensed information using Artificial Neural NetworksMultispectral Analysis