ML-Based Lightning Forecast

Internship Project Report

Submitted by

Mr. Deewakar Chakraborty
20-27-05

Under the kind guidance of

Mrs. Shikha Singh (Scientist 'C', IITM Pune)
Dr. Ramkrishna Bharath (Asst. Professor, Department of Applied Mathematics, DIAT, Pune)

Department of Applied Mathematics
Defence Institute of Advanced Technology
Girinagar, Next to Khadakwasla Dam, Pune,Maharashtra-411025
(Affiliated to Ministry of Defence)
Oct 2021

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Objective

Lightning casualties cause massive loss to life and property. Lightning is accountable for human injuries, the death of livestock and forest fires. It is also an effective source of electromagnetic interference that causes deterioration to electronic circuits, buildings, and other exposed man-made structures such as transmission lines, wind turbines and photovoltaics. This project focuses on building an AI-based lightning forecast model which can predict lightning occurrences in the near future. The forecast of lightning can be treated as a spatiotemporal sequence forecasting problem. The AI model will take meteorological features that affect the lightning and the sequence of past lightning occurrences as input and return future lightning occurrences as output.

1.2 Motivation

Lightning is a natural phenomenon that poses severe threats to human life, aviation and electrical infrastructures. Lightning forecast plays a crucial role in lightning disaster reduction. The existing forecast models are mainly based on numerical weather models and rely on lightning parameterization schemes. Meteorological experts manually build the lightning parameterization schemes. Unfortunately, these schemes can hardly benefit from abundant historical data. Unlike the standard prediction techniques that are totally based on numerical weather models, we propose a data-driven model based on neural networks for lightning prediction. The neural network has an encoder-decoder architecture. Firstly this forecast model extracts spatiotemporal features of simulations and observations via dual encoders, and then those extracted features are combined using a fusion module. Finally, fused features are given as input to the decoder to make forecasts.

Chapter 2

Literature Survey

In 2019, Yangli-ao Geng et al. [1] proposed a dual spatiotemporal encoder network model for lightning prediction over China known as LightNet. They suggested that existing lightning forecast models, based on numerical weather models, rely on lightning parameterization schemes have two significant drawbacks. Firstly, simulations of the numerical weather models usually have deviations in space and time domains, which introduces irreparable biases to subsequent parameterization processes. Secondly, the lightning parameterization schemes are designed manually by experts in meteorology, which means these schemes can hardly benefit from abundant historical data. Their paper proposed a data-driven model based on neural networks, referred to as LightNet, for lightning prediction. LightNet first extracts spatiotemporal features of the simulations and observations via dual encoders. These features are then combined by a fusion module. Finally, the fused features are fed into a spatiotemporal decoder to make forecasts.

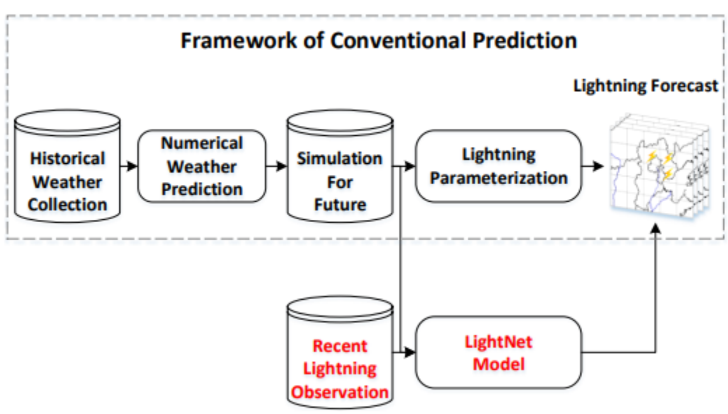

Figure 2.1: Proposed framework for lightning forecast in Yangli-ao Geng et al. [1]

Pradip Kumar Gautam et al. [3] published their work in 2021. Their paper conducted a study to determine significant meteorological predictors that determine lightning over the Indian subcontinent. In addition, they determined the correlation between various predictors and lightning occurrence. In their experiment, they have taken several meteorological variables to define the relative variable of lightning by using four models Linear Regression, Support Vector Machines(SVM), Gaussian process regression(GPR) and Regression Trees(RT). As a result, cape, k-index, RH, AOD, vertical velocity, Vertical integral of the divergence of cloud frozen of water flux and Vertical integral of divergence the cloud liquid of water flux showed good relation with lightning, which are defined based on the value of R and R2 score of different models.

The lightning prediction problem is a spatiotemporal sequence forecasting problem. Further, the survey is conducted for the paper Xingjian Shi et al. [2]. Their work proposed a convolutional LSTM (ConvLSTM) and used it to build an end-to-end trainable model for the precipitation nowcasting problem. Their work shows that the ConvLSTM network captures spatiotemporal correlations better and consistently outperforms FC-LSTM.

Chapter 3

Data Description

This study implemented the proposed forecasting scheme using data collected from meteorological stations in India between 2014 and 2020. The data contains information about the lightning activity such as the amplitude of lightning(in amp), latitude and longitude of the location where lightning has occurred and exact timing of the lightning occurrence. In addition, the other meteorological predictors used for forecastings, such as Cape, Relative humidity, the vertical integral of the divergence of cloud frozen water flux and cloud liquid water flux, were downloaded from the ERA-5 dataset. ERA5 provides hourly estimates of many atmospheric, land and oceanic climate variables.

3.1 Data analysis and visualization

The station data for lightning activity has three columns, the magnitude of lightning, location of its occurrence and time stamp associated with it. To analyze the trend of lightning, all of the lightning occurrences for each month are plotted over India's map. Figure 3.1 represents one such plot for lightning activity for January 2015.

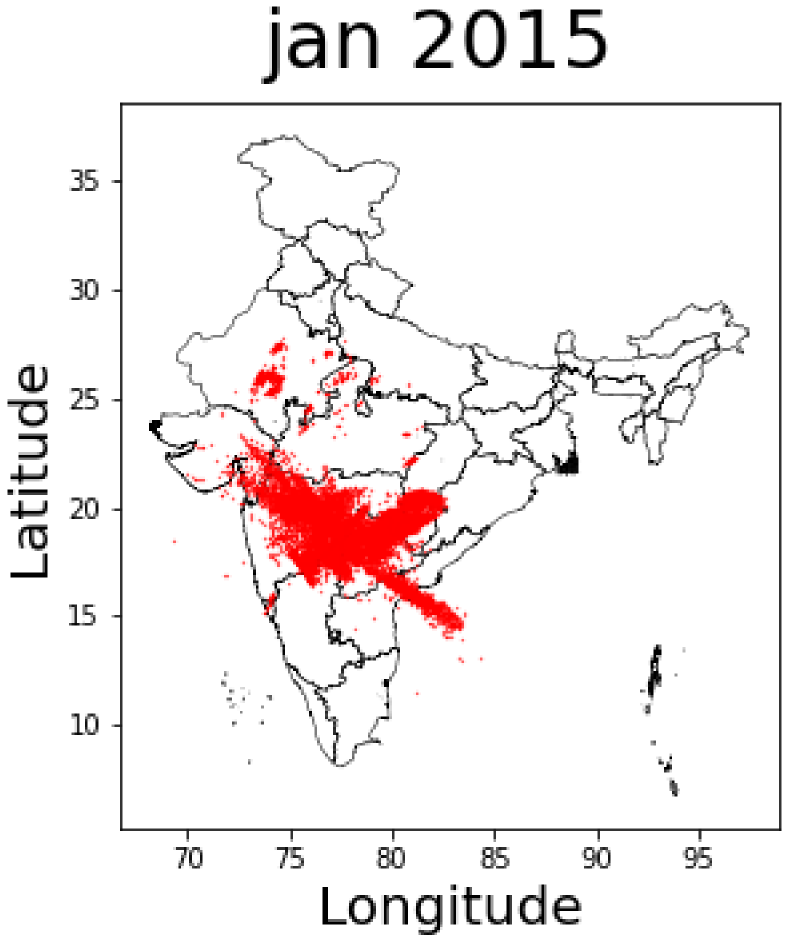

Figure 3.1: Plot showing lightning occurrences over India for Jan 2015

To analyze the trend for the magnitude of lightning, the maximum magnitude of lightning for each month is being plotted. Figure 3.2 represents one such plot for the year 2014.

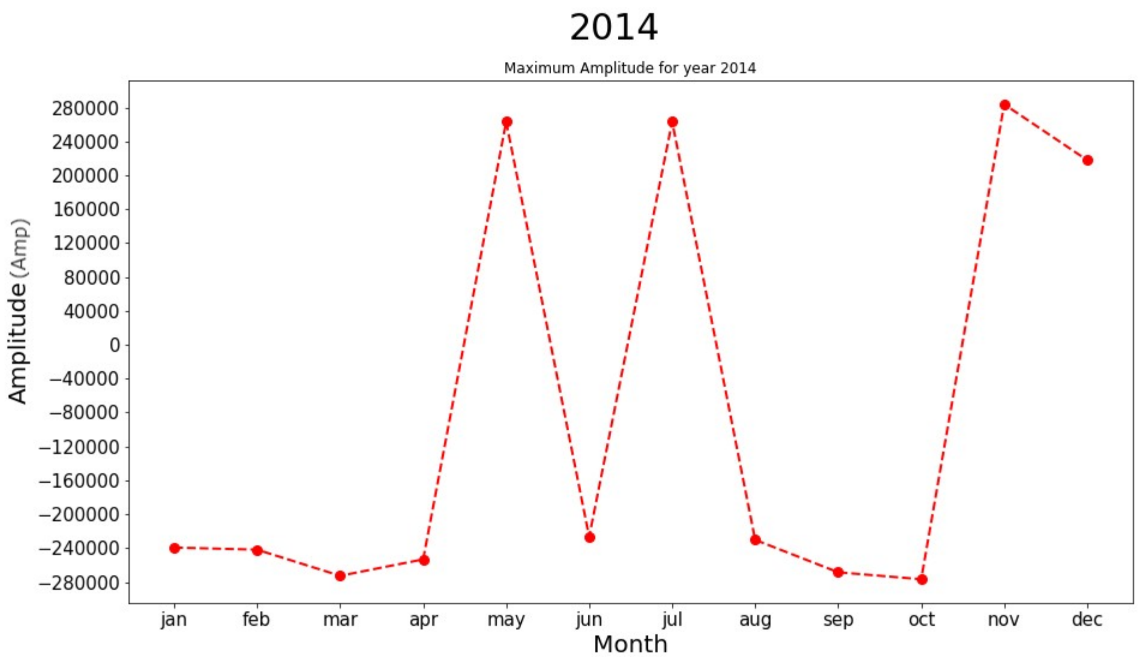

Figure 3.2: Plot showing maximum lightning amplitude observed for each month for the year 2014.

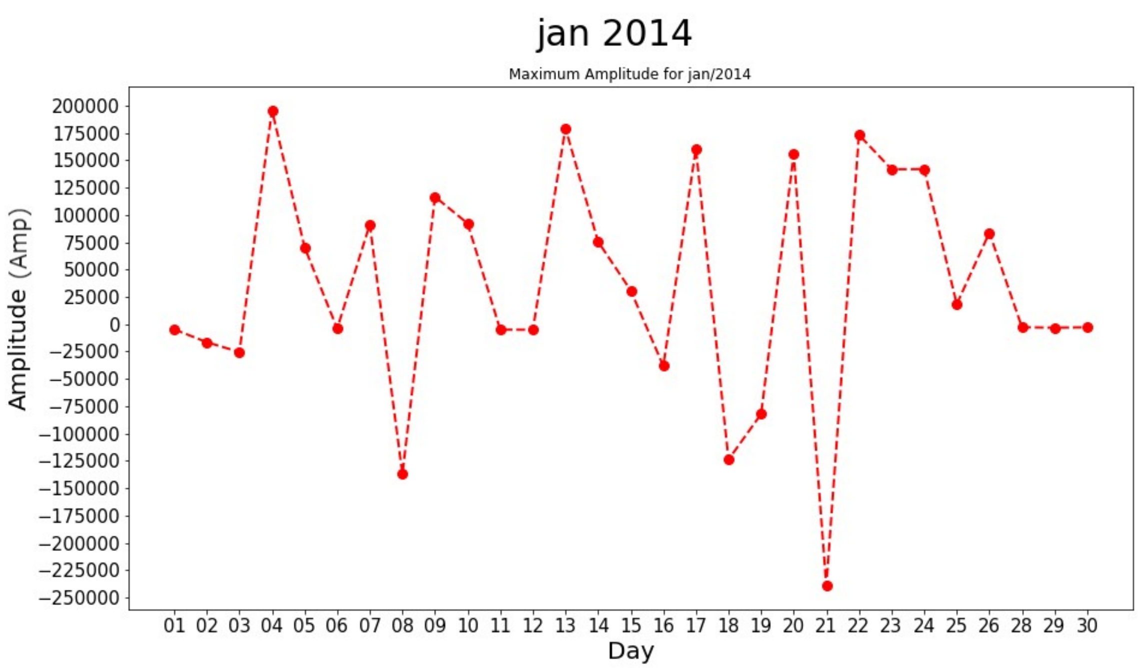

Figure 3.3: Plot showing maximum lightning amplitude observed for each day for January 2014.

Median plots give information about the median of lightning amplitude and its variance with maximum and minimum lightning magnitude.

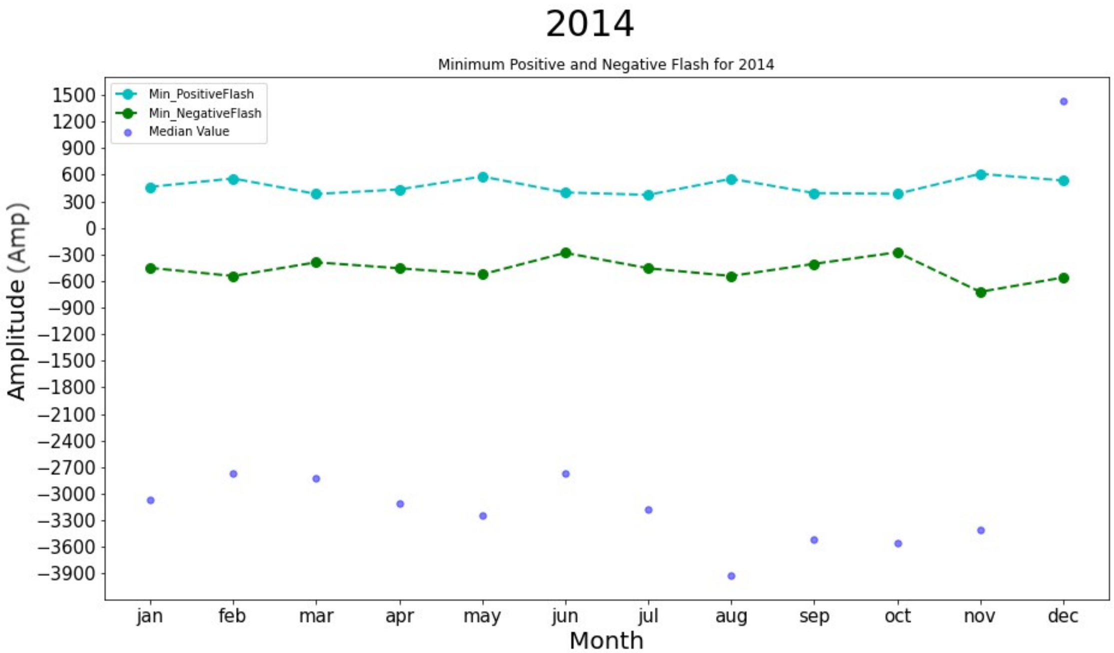

Figure 3.4: Sample median plot for year 2014

3.2 Data pre-processing

3.2.1 Data cleaning

The proposed lightning forecast scheme is currently focused on the region of Maharashtra. Therefore, the first step of data cleaning is to discard all the lightning occurrences outside the region of Maharashtra. Also, most of the lightning observed in Maharashtra is for April and May. Therefore, all the lightning activity apart from these two months are also discarded.

3.2.2 Feature engineering

The station data for lightning activity was in comma-separated file format(csv format). The csv format file was converted into the gridded format to simplify the forecast scheme. Gridded data is two-dimensional data representing a meteorological parameter along an evenly spaced matrix. The spatial resolution used is 2.5 degrees. This means each grid represents 277.5 kilometres. All the observed lightning activity is stored in appropriate grids according to the latitude and longitude of its occurrence. Also, timestamps of the lightning that are in continuous form is converted in 1 hour time intervals. Each time interval represents all the lightning activity that happened in that duration.

The occurrence of lightning has three dimensions now, latitude and longitude to represent the position and time of its occurrence is the third dimension.

Also, for simplification, the magnitude of lightning, a continuous value, is converted into a discrete format. Each point of lightning occurrence is now represented as a positive label, and the rest of the points are the negative labels.

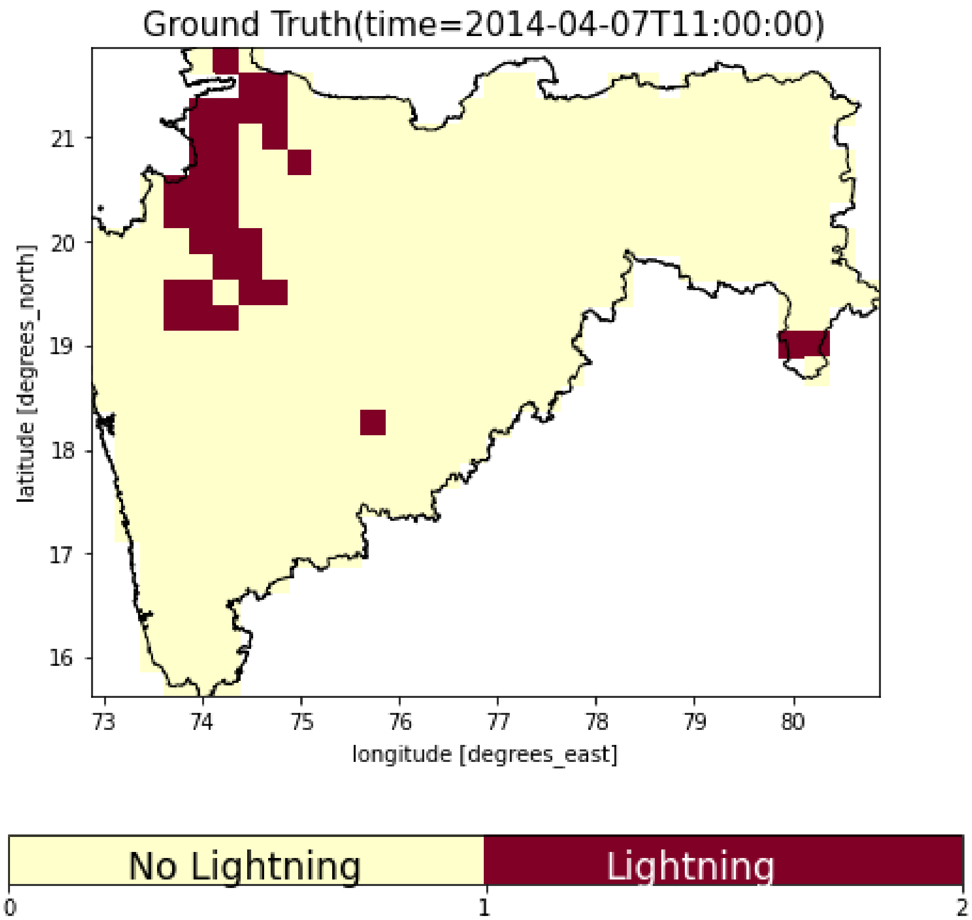

Figure 3.5: Plot showing lightning occurrence for region of Maharashtra

Along with the lightning activity the forecast model uses meteorological predictors. These predictors were downloaded from ERA-5 dataset and were already in gridded format with same spatial dimension of 2.5 degrees and temporal dimension of 1 hour.

Variable Details		
Variable	Dimension	unit
Convective available potential energy(CAPE)	10248x29x33	J kg**-1
cloud liquid water flux	10248x29x33	kg m**-2 s**-1
cloud frozen water flux	10248x29x33	kg m**-2 s**-1
Relative humidity	10248x500x29x33	%

Table 3.1: Table containing the details of variables

Chapter 4

Proposed Method

Lightning Nowcast is a spatiotemporal sequence forecasting problem with a sequence of past lightning occurrences and geographical features affecting it as input and a sequence of future lightning occurrences as output. However, such learning problems become non-trivial due to the high dimensionality of spatiotemporal sequences, especially when multi-step predictions have to be made unless the spatiotemporal structure of the data is captured well by the prediction model. RNN and LSTM models provide some insights on how to tackle such problems.

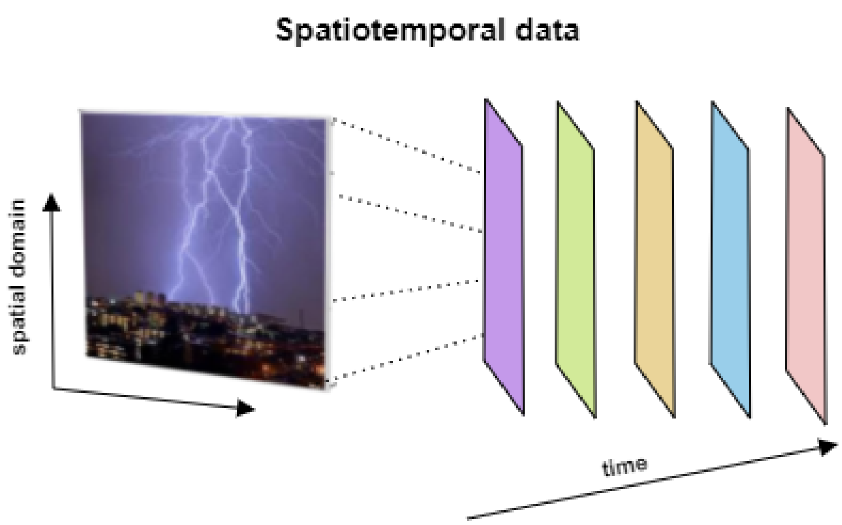

Figure 4.1: Sample sketch map of a spatiotemporal data

The proposed model solves this problem using Convolutional LSTM (ConvLSTM) layers. ConvLSTM layer in a neural network extends the idea of fully connected LSTMs by introducing convolutional filters in both the input-to-state and state-to-state transmission. By stacking multiple ConvLSTM layers and forming an encoding-decoding structure, we can build an end-to-end trainable model for a lightning forecast system.

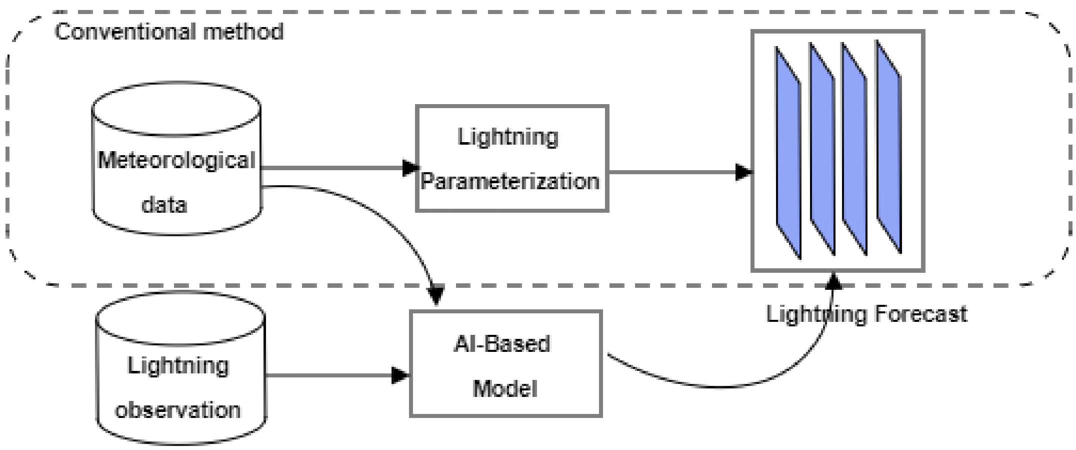

Figure 4.2: Block diagram for lightning forecast model

Figure 4.2 shows the framework of conventional lightning prediction method(inside the dotted box). Unlike conventional prediction, our AI-based forecast model introduces recent lightning observation data to assist prediction besides meteorological data.

4.1 Model in study

Figure 4.3 represents the overall architecture of the model. The model comprises of an encoder, fusion module and the prediction decoder. Since the meteorological data(for last six hours) and the lightning observation(for last three hours) belong to different time domains there are two encoders. Both the encoders convert the raw data into feature codes in a unified space for further combination. Then output of both the encoders is fused using fusion module. Finally, the prediction decoder decodes the lightning prediction using these fused features.

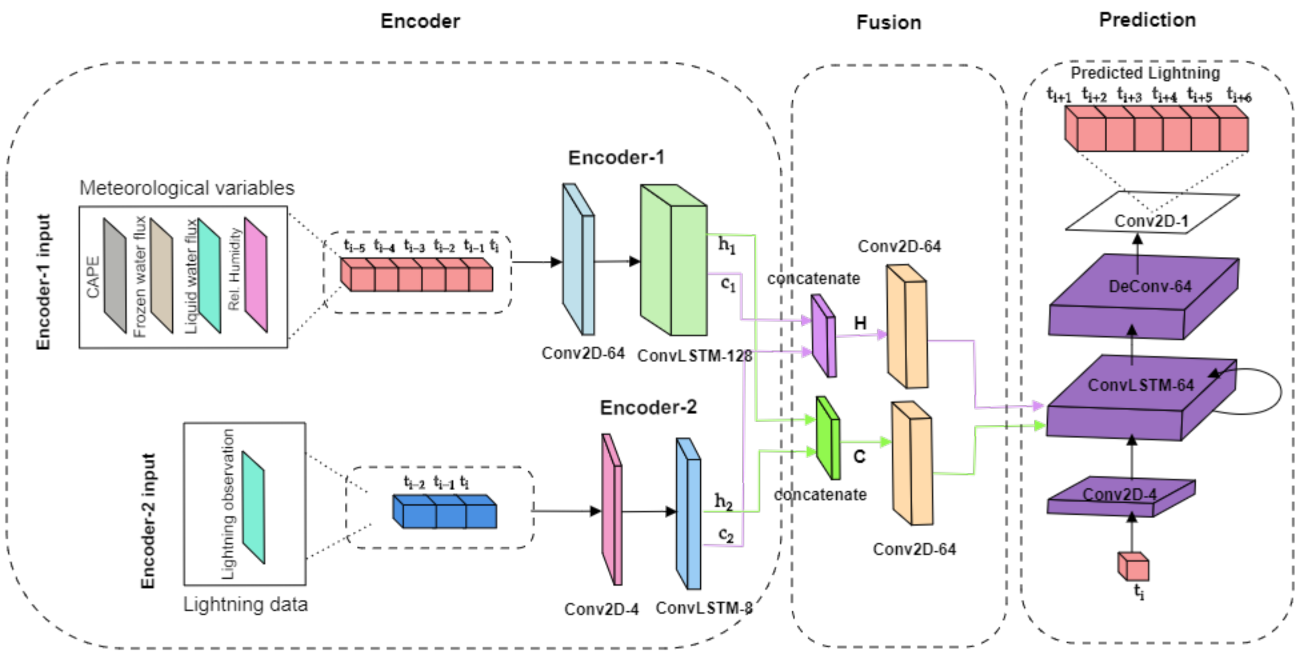

Figure 4.3: Detailed architecture of the model

Module	Notation	Size	stride
WRF Encoder	Conv_1	$[7 \times 7, 64]$	2
	ConvLSTM	$[5 \times 5, 128]$	1
Obs. Encoder	Conv_2	$[7 \times 7, 4]$	2
	ConvLSTM	$[5 \times 5, 8]$	1
Fusion Module	Conv_3	$[1 \times 1, 64]$	1
	Conv_4	$[1 \times 1, 64]$	1
Pred. Decoder	Conv_5	$[7 \times 7, 4]$	2
	ConvLSTM	$[5 \times 5, 64]$	1
	DeConv	$[7 \times 7, 64]$	2
	Conv_6	$[1 \times 1, 1]$	1

Table 4.1: Detailed settings of the architecture of neural net

4.2 Initial results

The proposed lightning forecast model is initially trained for six hours lead time, the input time step for the meteorological data and lightning observation is six and three hours, respectively. Data from 2014 to 2017 are treated as training data, data from 2018 as validation data and 2019 to 2020 as testing data. Metrics to evaluate the training process are the probability of detection (POD) and false alarm ratio (FAR). Let N_1, N_2, N_3 and N_4 denote the number of true-positive, false-positive, false-negative and true-negative grids, respectively. The equations for the three metrics are detailed in the table 4.2.

Metrics Details			
Name	Equation	Range	Explanation
POD	$\frac{N_1}{N_1+N_3}$	[0,1]	The ratio of the number of hit lightnings to the number of observed lightnings.
FAR	$\frac{N_2}{N_1+N_2}$	[0,1]	The ratio of the number of false alarm lightnings to the number of foretasted lightnings.

Table 4.2: Detailed information of the metrics used

Due to the imbalanced nature of data, a weighted loss function is used. Weight in the weighted loss function is four which is picked heuristically and needs to be hyper tuned and experimented with in the future. Adam optimizer is used for training the model.

Initially, the model is trained for 100 epochs only as it starts to overfit the training data after that. Next, the model is trained using Nvidia's Tesla P100 accelerator with a memory of 12 GB.

Metric	Performance
POD	0.3796
FAR	0.69422

Table 4.3: Model performance over test data after 100 epoches

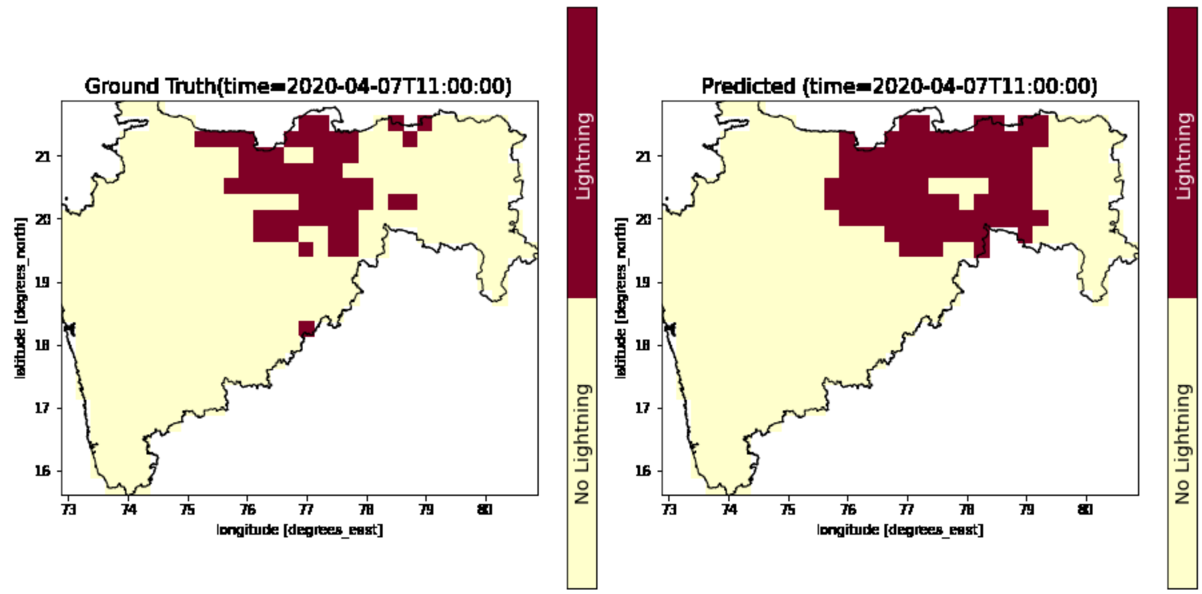

Figure 4.4: Sample lightning forecast for 11 am, April 7, 2020.

4.3 Future work

After getting satisfactory initial results, future steps include the addition of more meteorological predictors that affect the lightning over Maharashtra, which may improve the model's performance. Also, tuning of hyperparameters is required. Initial hyperparameter values are heuristic. Experimenting with the hyperparameters can also result in improvement of the model performance. The final step will compare our model's results with the existing baselines. We have an existing numerical model for lightning forecast in IITM Pune for baseline. The purpose of this AI-based forecast model is to perform better than the baseline model.

Bibliography

[1] Geng, Yangli-Ao Li, Qingyong Lin, Tianyang Jiang, Lei Xu, Liangtao Zheng, Dong Yao, Wen Lyu, Weitao Zhang, Yijun. (2019). LightNet: A Dual Spatiotemporal Encoder Network Model for Lightning Prediction. 2439-2447. 10.1145/3292500.3330717.

[2] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, Wang-chun Woo. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, arXiv:1506.04214, September 2015.

[3] Pradip Kumar Gautam, Deweshvar Singh. Defined the predictors of the lightning over India by using artificial neural network arXiv:2104.13958, April 2021

[4] Mostajabi, A., Finney, D.L., Rubinstein, M. et al. Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. npj Clim Atmos Sci 2, 41 (2019). https://doi.org/10.1038/s41612-019-0098-0

[5] Litta A. J, Sumam Mary Idicula and C. Naveen Francis 2012. Artificial Neural Network Model for the Prediction of Thunderstorms over Kolkata. International Journal of ComputerApplications (0975 – 8887). Volume 50 – No.11, July 2012.

[6] Ying Wang, Yi Yang, and Shuanglong Jin. 2018. Evaluation of lightning forecasting based on one lightning parameterization scheme and two diagnostic methods. Atmosphere 9, 3 (2018), 99.

[7] Omvir Singh, and Jagdeep Singh, Lightning fatalities over India: 1979–2011. Meteorol. Appl. 22: 770–778 (2015).