Methodology for the BTP Project :

Task-1 : Correlation between LULC and NightLights

- We have taken Night Lights Data from VIIRS satellite : Link
- We have taken LULC Data from MODIS satellite : Link
- The region of interest is Washington. (Because, we were able to see good amount of Night Lights there)

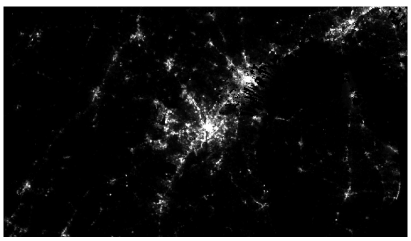
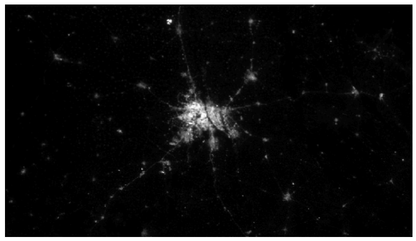
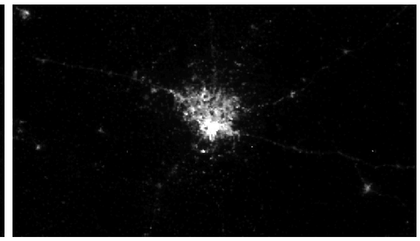
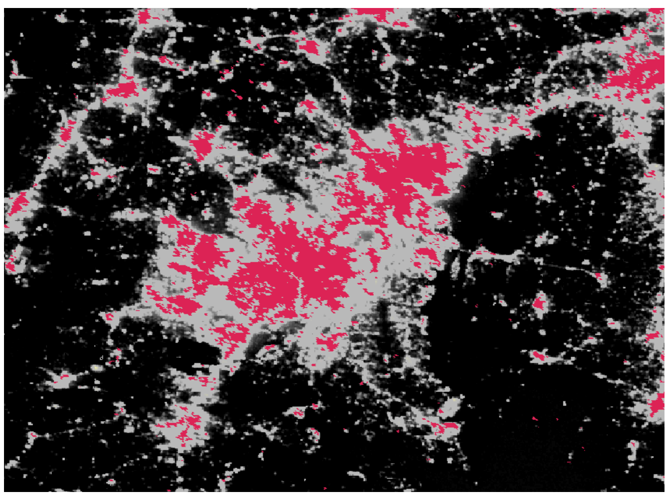




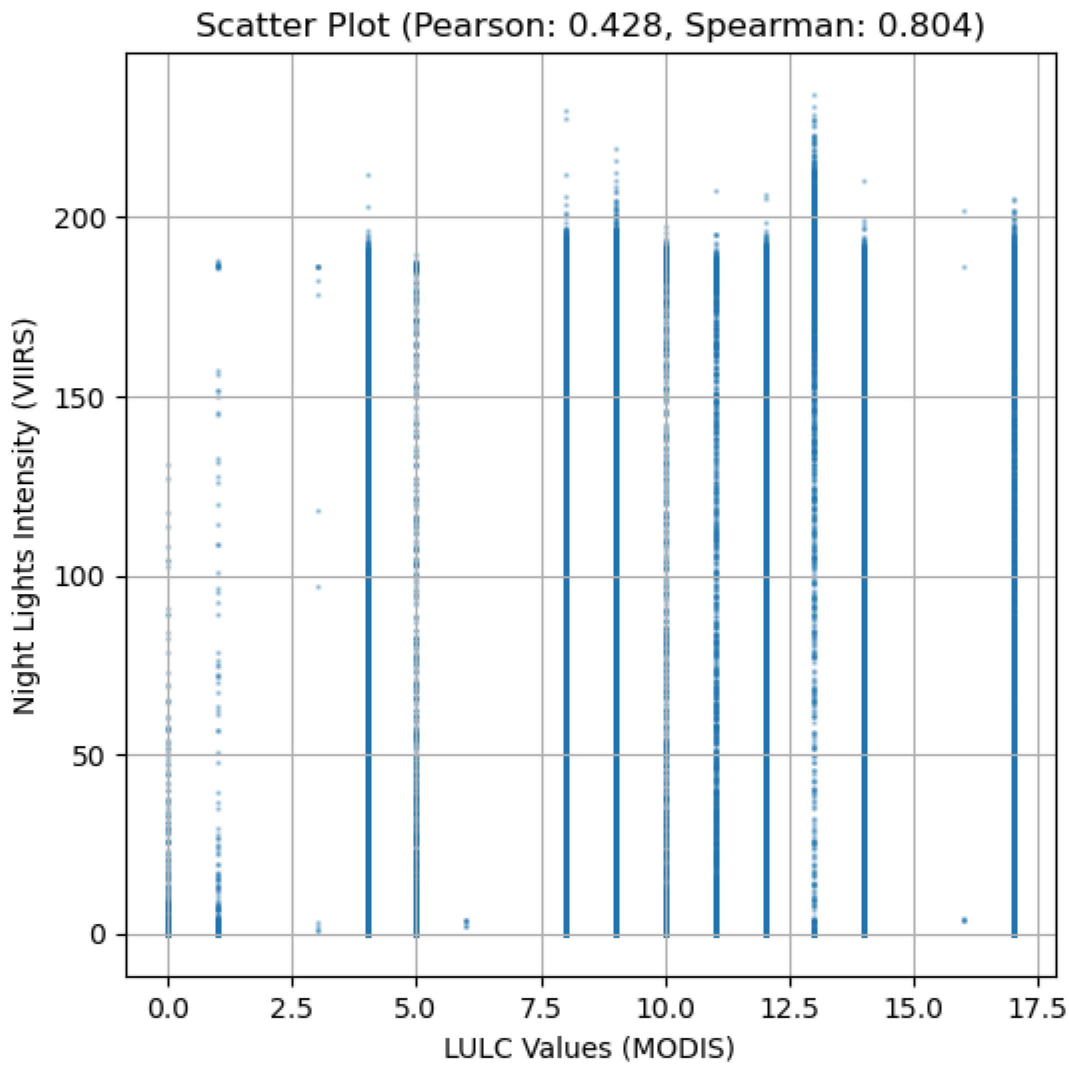
Fig : Washington Fig : Delhi Fig : Hyderabad

- The VIIRS data was found in ESPG coordinates and MODIS was found in Sinusoidal. So we converted MODIS to ESPG. But there we still some errors (Map Shifts). So we manually shifted the map. : Link
- Then we overlaid One map on another. :



Red colour is the Land Cover, White / Grey Colour is the Lighted Region.

- We found the following score correlation between night lights and LULC



1. Pearson Correlation: 0.428
    a. Indicates a moderate linear relationship, but with significant variability.
    b. Suggests that LULC values alone do not strongly predict night light intensity in a strictly linear manner
2. Spearman Correlation: 0.804
    a. Shows a strong rank-based correlation, meaning night light intensity generally increases with LULC values, though not in a linear fashion.
    b. Suggests that LULC categories influence night light intensity in a structured but non-linear way.

Observations from Scatter Plot:

a. LULC values are discrete (categorical), forming vertical bands.

b. Higher night light intensity corresponds to specific LULC classes (likely urban areas).

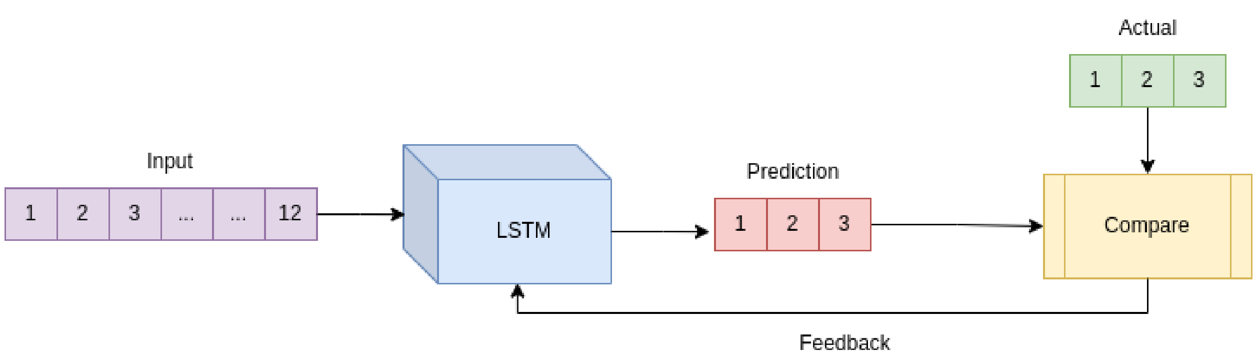
c. Some LULC classes (e.g., forests, water bodies) have very low night light values.

LC_Type1 Class Table

Value	Color	Description
1	#05450a	Evergreen Needleleaf Forests: dominated by evergreen conifer trees (canopy >2m). Tree cover >60%.
2	#086a10	Evergreen Broadleaf Forests: dominated by evergreen broadleaf and palmate trees (canopy >2m). Tree cover >60%.
3	#54a708	Deciduous Needleleaf Forests: dominated by deciduous needleleaf (larch) trees (canopy >2m). Tree cover >60%.
4	#78d203	Deciduous Broadleaf Forests: dominated by deciduous broadleaf trees (canopy >2m). Tree cover >60%.
5	#009900	Mixed Forests: dominated by neither deciduous nor evergreen (40-60% of each) tree type (canopy >2m). Tree cover >60%.
6	#c6b044	Closed Shrublands: dominated by woody perennials (1-2m height) >60% cover.
7	#dcd159	Open Shrublands: dominated by woody perennials (1-2m height) 10-60% cover.
8	#dade48	Woody Savannas: tree cover 30-60% (canopy >2m).
9	#fbff13	Savannas: tree cover 10-30% (canopy >2m).
10	#b6ff05	Grasslands: dominated by herbaceous annuals (<2m).
11	#27ff87	Permanent Wetlands: permanently inundated lands with 30-60% water cover and >10% vegetated cover.
12	#c24f44	Croplands: at least 60% of area is cultivated cropland.
13	#a5a5a5	Urban and Built-up Lands: at least 30% impervious surface area including building materials, asphalt and vehicles.
14	#ff6d4c	Cropland/Natural Vegetation Mosaics: mosaics of small-scale cultivation 40-60% with natural tree, shrub, or herbaceous vegetation.
15	#69fff8	Permanent Snow and Ice: at least 60% of area is covered by snow and ice for at least 10 months of the year.
16	#f9ffa4	Barren: at least 60% of area is non-vegetated barren (sand, rock, soil) areas with less than 10% vegetation.
17	#1c0dff	Water Bodies: at least 60% of area is covered by permanent water bodies.

Task - 2 : Future LULC Prediction using ML :

- We downloaded LULC data of Hyderabad region from 2016 to 2024 each month from sentinel satellite : Link Dataset_Link
- We were unable to extract entire data at a time. So we divided the map into 4 tiles and downloaded each tile. While processing, we merged all the four tiles into a single map.
- We converted the maps (tiff files) into numpy arrays for further ML operations.
- We Handled missing data using temporal interpolation:
  - Replaces zeros or invalid values with NaN
  - Uses forward fill to fill missing values with the last valid observation
  - Uses backward fill for any remaining NaNs at the start of the time series
- The ML model which we are using is LSTM (Long Short-Term Memory). Which follows neural network architecture.
  - We have divided the data as :
    - X = 12 months LULC data
    - Y = 3 months LULC data (after X)
  - Thus each time we train with the above X and Y Supervised learning and update the weights of Neural Networks (LSTM)



- We are using LSTM because it remembers the previous data and also it is faster in computation.
- Till now we have completed preprocessing of the data [converting into numpy arrays].
- We now have the build the model and train it with the data.
- We were facing insufficiency of GPUs, so it's taking a lot of time to debug and train the model.