Mapping Lightning Strikes across Gujarat State for the Year 2022 and Development of Prediction Model

Pratyush Rao[1], Gautam Shah[1], Nimit Savant[1], Seema Lakhani[1], Alok Taori[2*] and Pallavi Ghalsasi[1,*]

[1]School of Engineering and Technology, Navrachana University, Vadodara, Gujarat, India
[2] Atmospheric Science Division, Earth & Climate Science Area, National Remote Sensing Center (NRSC), Indian Space Research Organization (ISRO), Hyderabad, India
Email: pallavig@nuv.ac.in; alok_taori@nrsc.gov.in

(Received on 27 December 2024; in final form on 14 April 2025)

DOI: *https://doi.org/10.58825/jog.2025.19.1.208*

Abstract: Under the influence of changing climate, extreme events are getting more frequent. Atmospheric lightning is considered a natural disaster which accounts for more than 40% of the weather-related deaths worldwide. Unlike the other weather-related phenomena, mitigation efforts pertaining to lightning need high resolution information on the vulnerable regions with region specific prediction. Present study is focused on mapping lightning incidences in the state of Gujarat, India, for the year 2022 and determines the vulnerable locations within a state. Our studies have shown that during monsoon season, most of the lightning strikes happen where the elevation is less than 30 meters above sea level and post monsoon the built area poses a serious concern. Vulnerability of strikes also increases when the temperatures are between 21-30 ^{0}C and curling wind speed is between 0-2 m/s. Our prediction model could predict 87% records correctly. The extended objective of the study is to develop a prediction model that aims to help implementing the appropriate nationwide mitigation efforts in the vulnerable geographical regions.

Keywords: Lightning, lightning detection sensor, vulnerable locations, mapping, elevation, land coverage, weather parameters, prediction model

1. Introduction

Atmospheric lightning is an important natural disaster known to mankind for a very long time. It is understood that Lightning is a sudden, electrostatic discharge creating a giant spark or flash which occurs when charged atmospheric regions temporarily equalize themselves resulting in this high energy phenomena (Rakov and Uman 2003). Lightning occurs due to convective processes which is a combination of humid air lifting up in the troposphere under the influence of vertical winds. Owing to this, though lightning occurs in all seasons, it is noted that during monsoon season when thunderstorms are generated frequently, the lightning occurrence is maximum (Rocque et al. 2024).

For normal atmosphere, air acts as a dielectric- a material that does not conduct electricity easily and is used to separate conductive materials, between the opposite charges within the cloud, between different clouds and between the cloud and the ground. Upon sufficient buildup of the opposite charges, dielectric breakdown happens and there is a rapid discharge of electricity. The reason for the charge build up is yet to be known completely but is thought to be a turboelectric effect in which electricity is generated by friction. The heavy graupel particles (precipitation consisting of snowflakes) coated with supercooled water droplets (liquid form below its normal freezing point due to a lack of nucleation site for ice crystal formation) moving downwards gain negative charge and light ice flakes along with small, supercooled water droplets moving upwards gain positive charge during their collisions. Depending on the location of charge build up, lightning can occur as intra-cloud lightning, cloud to cloud lightning or cloud to ground lightning (Rakov & Uman 2003). This process holds true at low temperatures and not at higher temperatures. This temperature dependence results in the typical tripole structure with a central negative charge region between 0° and -10°C.

The polar distribution of charges along with movement of charges are responsible factors for the spread and branches of lightning discharge in space and time resulting in multiple lightning patterns like bead lightning, ribbon lightning and anvil crawlers etc.

As the world is becoming increasingly susceptible to rapidly changing climatic conditions, it is also using advanced technology (Cummins et al. 1998) to mitigate the loss of life and damage to property during such incidents and build a robust disaster management system. This becomes further important because studies have shown that with rise in the tropospheric temperatures, the lightning occurrences are bound to increase (Romps et al., 2014). Akin to the other parts of the world, lightning over India is fatal with a claim of more than 40% deaths occurring because of natural phenomena (National Crime Record Bureau reports 2002-2022). To monitor lightning occurrences over India, National Remote Sensing Centre (NRSC), ISRO has established a network of long-range lightning detection sensors (LDS) with each sensors having a separation of about 300 km, to gather the data on cloud-to-ground (CG) lightning occurrences (Taori et al. 2022). Before the present study, this LDS network data has been compared with other co-existing lightning detection network in India (Taori et al., 2022) and world-wide lightning location network (Taori et al., 2024) and to study the diurnal characteristics of lightning occurrences over India (Taori et al., 2023).

Over the years, many researchers have analyzed lightning for its global, seasonal, and regional distribution, their correlation with topography, land cover and weather parameters and influence of urbanization and climate change, prediction models.

Ghosh et al. 2023 studied four years of lightning observation data to determine the seasonal and spatial distribution of the ratio of intra-cloud (IC) lightning to cloud-to-ground (CG) lightning in thunderstorms over the Indian subcontinent and they found that the ratio is high (6–10) in the northwestern parts and low (0.5–3.5) in the northeastern parts. Their study also suggested the importance of separating CG lightning flash from total and can be used in numerical models to give a proper prediction of CG lightning in hazard mitigation. Research work by Holle et al. 2016 analyzes lightning data over the contiguous United States (CONUS). This work along with earlier study of the diurnal variations of CG flashes over the CONUS by Holle et al. 2014, makes it possible to identify time and location of the lightning occurrence over the CONUS and adjacent regions by season, month, week, and time of day and thereby help in lightning related disaster mitigation. Seasonal variation of lightning activity over the Indian subcontinent by Ranalkar et al. 2009, investigating lightning patterns over India, found a strong correlation between lightning activity and the Indian monsoon season, with a significant increase in lightning strikes during this period. Research of Hazmi et al. 2023 conducted statistical analysis about the correlation between lightning and several weather parameters (rainfall, humidity, air temperature, and wind) in Padang from 2016 to 2020. These results indicated a strong relationship between lightning and rainfall during the wet season and rainfall being the only primary variable in lightning frequency. The research study by Williams et al. 1992 and Price et al. 1992 utilized satellite data to map global lightning distribution and it confirmed that tropical regions experience the most lightning activity due to favorable conditions for frequent thunderstorms and lightning occurrence decreases as we move away from the equator towards the poles. Cecil et al. 2014 observed gridded climatologies of total lightning flash rates using spaceborne optical transient detector (OTP) and lightning imaging sensor (LIS). According to their work, central Africa showed the greatest yearly average of lightning strikes, and the mean global flash rate is 46 flashes/s varying between 36 in February to 60 flashes/s in August (Cecil et al. 2014). Holzworth et al. 2019 studied distribution of super bolts having energy > 1MJ which peaks globally in the northern hemisphere during winter (November-February) with most of them in western Europe, mediterranean sea and strong local maximum over the Andes in south America with fewer bolts in the north pacific east of Japan. A very few bolts were found during April-October while the local occurrence peaks do not coincide with the usual 3 main lighting chimneys and the super bolts spread predominantly over oceans and seas (Holzworth et al. 2019).

Spatial patterns and land surface features associated with Cloud to Ground lightning in Bangladesh were explored by Dewan et al. 2022 wherein they found a negative correlation with vegetation cover, suggesting a preference for open areas. Studies on distribution of lightning in relation to topography and vegetation cover over the dry and moist regions in the Himalayas by Oulkar et al. 2019 have shown that lightning activity is more at a higher terrain slope/elevation in the dry NW region where vegetation cover is less, whereas it is more at a lower terrain slope/elevation in the moist NE region where vegetation cover is more. The relation of lightning with elevation and vegetation cover has been studied in the Mediterranean (Kotroni et al. 2008) and in Shandong region of China (Wu et al. 2020).

In 2018, Yair et al. (2018) reviewed lightning hazards to human societies in a changing climate reviewed increasing risk of lightning hazards due to population, urbanization, and climate changes. An online document by Lee et al. (2011) mentions that there is a misconception that only taller objects attract lightning as compared to shorter objects. In fact, the objects on the ground have little influence on the travelling path of lightning until the stroke comes to a distance roughly equal to the height of the object.

Recent development in the area of Statistical models and Machine Learning algorithms (Singh et al, 2021, Wang et al. 2023, Raheem et al. 2023) show that they can analyze historical weather data along with lightning occurrences to identify patterns, understand complex relationships between causative weather variables and lightning which can then be used to predict the lightning occurrences in future weather scenarios. However, these methods require high quality large datasets and may not generalize to unseen weather patterns.
Apart from these there are researchers studying the influence of urban heat islands on lightning patterns and synoptic background for barrier effect over built-up areas, effect of aerosols etc. (Shi et al. 2023, Shi et al. 2024). Researchers are also studying the variation of lightning ignited wildfire patterns under climate change (Perez-Invernon et al. 2023).

While global trends for lightning patterns are getting established, a deeper understanding of how local factors like atmospheric circulation patterns influence regional lightning variations is needed. With this in view, present study maps the lightning incidences in the state of Gujarat, India, for the year 2022. We have analyzed the occurrences with respect to weather parameters and various land use land cover parameters. Previous studies often focus on specific regions or neglect detailed mapping and prediction models across larger geographical areas. The present study specifically targets the entire state of Gujarat, offering a detailed analysis of lightning incidents across varied landscapes including water bodies, forests, and urban areas. This study advances the field by developing a machine learning-based prediction model using XGBoost algorithm (XGBoost basics (2024); KDE Estimation, PCA Basics). The model aims to forecast lightning strikes based on environmental variables, thereby providing feature selection using machine learning and assisting in the implementation of preventive measures and disaster management strategies. Furthermore, it utilizes QGIS software for mapping and spatial analysis, coupled with integration of diverse datasets (e.g., elevation data, weather parameters) that enhances the accuracy and applicability of findings. Thus, our work on prediction model aims to foresee the CG lightning events and help the Government implement protective policies over the vulnerable geographical regions.

2. Data collection

Lightning Data Collection

Lightning detection sensor (LDS) installed by National Remote Sensing Center (NRSC), Indian Space Research Organization (ISRO), Hyderabad confirms the geographic location of a lightning strike using time of arrival (TOA) algorithm (Taori et al. 2021). In this method, all stations are assumed to be synchronized by time. When lightning occurs, a constant difference in arrival time is noted between two stations defining a hyperbola. Multiple stations provide multiple hyperbolae and the intersection of those defines the location. When at least 4 sensors detect this discharge, the location is mapped for lightning. LDS can capture lightning strikes within a radius of 300 km with more than 98% accuracy. The wide network of LDS has given the authors access to the data for the entire state of Gujarat through Bhuvan web portal (Bhuvan) available at https://bhuvan-app1.nrsc.gov.in/lightning/.

Weather Data Collection

Lightning was correlated with weather dependent factors (variable) like wind, temperature, and geographic factors (fixed) like elevation, ground coverage etc. The data pertaining to these factors was carefully selected and obtained from reputable repositories (as given below) in raster or GRIB format. Elevation data in raster format was obtained from (Diva-gis) https://www.diva-gis.org/gdata . Ground Coverage Data (ArcGIS) in raster format was obtained from https://livingatlas.arcgis.com/landcover/. The weather data related to wind and temperature in GRIB format (Copernicus, NOAA) was obtained from https://cds.climate.copernicus.eu/. Weather data in number format required for the calculation of prediction model was downloaded from (Open-Meteo) https://open-meteo.com.

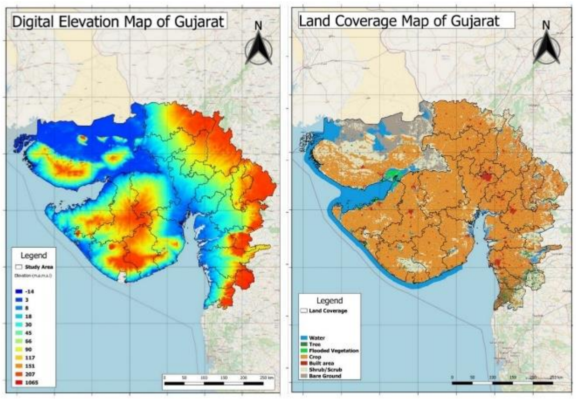
3. Methodology

Mapping of Lightning Data

Lightning data thus collected have been mapped on the state map of Gujarat using freely available QGIS software. Gujarat state district shape file was used as a base map and was overlaid by lightning data containing the geographical coordinates. The default option in QGIS is to make sure that each overlaid map is projected and aligned correctly without having to use coordinate reference system (CRS) option in QGIS. The lightning map was overlaid with a certain percentage of transparency so that the base map of districts could show through as a reference map. For better visualization of the data from GRIB file, display properties of the layer were changed for colors in the map, data range etc.

The spatial distribution of lightning strikes was compared with elevation, land cover, and weather parameters (wind speed, wind direction, and temperature) through visual analysis of the mapped data.
the elevation raster (Figure 1a) and land cover raster (Figure 1b) was overlaid with The monthly aggregated lightning strike maps (Figure 2) using QGIS software as a heat map, with darker shades representing higher concentrations of strikes in a given area. This allowed for assessing the relationship between strike locations and factors such as elevation and land cover types (water bodies, built areas, forests). Additionally, for the day with maximum lightning strikes in each month, the strike distribution was visually compared with the simulated wind speed, wind direction (Figure 3), and temperature (Figure 4) data obtained from GRIB files. These weather parameters were mapped using QGIS, and the spatial patterns were analyzed in relation to the observed lightning strikes.

Furthermore, specific cases were investigated where the effect of elevation on strike distribution is prominent, such as the analysis for August 31, 2022 (Figure S1). The elevation raster and wind direction data were mapped, and the strike locations were analyzed concerning high-elevation areas and wind flow patterns.



(a) (b)

Figure 1. Elevation map (a) and land cover map (b) for Gujarat State

Data Handling

In Gujarat, cloud to ground lightning strikes are less frequent phenomena compared to the central and east coastal states of India (ref. Taori et al., 2022). The dataset engineered to use for this prediction model, consists of date, co-ordinates of a region, minimum and maximum temperature on a given date, minimum and maximum apparent temperature on the said date, wind speed, total rain received, total number of hours it precipitated for, shortwave radiation and count of lightning strikes.

Most of the records in the dataset have 0 lightning strikes. Developing a regression model that can give the exact number of lightning strikes per day will be difficult for such a dataset because of that very reason. The approach of prediction used in this dataset is known as a classification model. Instead of predicting the exact

number of lightning strikes we have tried to predict the density of lightning strikes (number of lightning strikes per 10 km x 10 km pixel) that the region will be receiving.
To achieve this, the lightning strike count column is divided into two bins. We assign each bin a number for the classification model to interpret. In the model, 0 translates to 0 expected lightning strikes, 1 translates to more than one lightning strikes. It should be noted that this is the methodology that worked the most effectively.

To test our predictions, the dataset has been divided into two halves, training and testing set. The data in the training set has been trained and then the same model is used for predicting the values on the testing set for comparison of the predictions to the actual values. The evaluation of our model is based on these predictions. The number of lightning strikes in the dataset is mostly zero, in order to develop a model to be more sensitive towards detecting the outlier's majority of records that have zero lightning strikes are deleted known as under-sampling.

Since the dataset in the present study contains more than 3 different weather variables like temperature, rainfall, shortwave radiation, wind speed, wind direction, snowfall, precipitation and evapotranspiration, which are hard to visualize in a Cartesian co-ordinate system, it is compressed for visualization purposes using Principal component analysis (PCA). PCA is a dimensionality reduction technique used in linear algebra, which transforms the data into a new set of reduced number of orthogonal variables called principal components. These components are linear combinations of original variables which can explain the variance of all the variables to the maximum. Detailed information is given in the Supporting Information (PCA).

For our prediction model, the classification report has been generated using a machine learning classification model known as XGBoost, which is a tree-based algorithm that uses the attributes provided in the dataset to classify an instance into either a record with lightning strike or a record without lightning strike without excessive scaling. More details on this algorithm can be found in Supporting Information (XGBoost).

4. Results and Discussions

Mapping of lightning incidents with land coverage and simulated weather data

To comment on whether the said factors do affect lightning incidents or not and their significant effects on the frequency and distribution of the strikes, the GRIB/Raster files for each factor (Elevation (Figure 1a), land coverage (Figure 1b), wind speed and direction (Figure 3) and temperature (Figure 4), were compared to the monthly aggregated level incidents and on maximum day respectively.

From the data (Figure 2), it is observed that February, May, November, and December received almost no lightning strikes whereas during June, July, and August almost all regions in the state received lightning strikes. The month of September received maximum lightning strikes. This may be likely due to convective activity which is the mother cause of lightning while rainfall can be associated with local convective rain (Singh et al., 2014, Sharma et al. 2020) or the large-scale monsoon circulation.

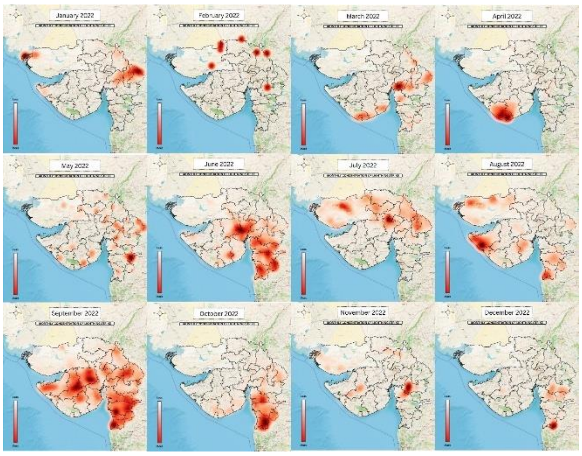


Figure 2. Month wise aggregated lightning strikes from January to December 2022

The most vulnerable locations shifted from northeast to southwest from January to April 2022. For the months of May and June, these vulnerable locations shifted from southwest to southeast and northeast. For the months July and August, the vulnerable locations remained in southeast locations and shifted from northeast to northwest. Finally, during September to December, the vulnerable locations remained in the southeast.

It is normally expected that geographic locations with high elevation can receive more lightning strikes as these regions can provide the shortest path for the equalizing charges to complete the circuit (Oulkar et al. 2019, Kotroni et al. 2008, Wu et al. 2020). Lightning strikes are typically observed in regions of medium to high elevation (30–1065 meters above mean sea level), with the exception of July–September. The majority of lightning strikes during these monsoon months occur at medium to low elevations (less than 30 meters above mean sea level).

This may be due to the reason that during monsoon the multilayer clouds are more frequent (Kundu et al. 2023), and the vertical extent is larger (Sharma et al. 2022), which may overcome the minimum path of lightning occurrences owing to the elevation effect. Thus, inference on elevation may be difficult as the effect of elevation on lightning is season dependent and it holds a very good relation in pre-monsoon seasons.

There happens to be another incident where significant effect of elevation was seen, on 31st August 2023 (Figure S1) (the day maximum lightning strikes occurred in August) winds were flowing from west direction to east. These winds could not flow past the highly elevated regions in south of Gujarat and hence the regions right before that experienced an insane number of occurrences. The same thing happened in Dwarka region of Gujarat in the month of August, it was the only time this region had an enormous count of strikes which is possibly due to the very high elevation of the region right past Dwarka. Many such cases from September to December were observed where there are an exceedingly high number of strikes right before the regions with very high elevation of land. With the help of models, it was shown that depending on the season, the effect of regional ground altitude on lightning striking is important. Chen et al. (2012) noted that the striking distance increases significantly as the ground altitude increases. This is found to be because the critical electric field necessary for sustaining the negative streamer zone decreases as the ground altitude increases. Hence, it may be inferred that high elevation affects the frequency and distribution of lightning incidents.

The land coverage map of Gujarat has several types of regions like water bodies, built areas and forest regions. Interestingly in monsoon season especially in August, regions with proximity to water bodies and forest received more strikes than the urban spread. Water bodies make an easy target because of the salt concentration which provides them with conducting properties and the trees in forests provide much needed elevation and pointed objects (Asfur et al. 2020, Oulkar et al. 2019, Kotroni et al. 2008, Wu et al. 2020). However, the focus has also been on the built area from September till November. One explainable reason as to why built areas have so many incidents is that the frequency of lightning strikes is directly proportional to the urbanization caused by the enhanced pollution (Lal & Pawar 2011). This, together with high-rise establishments which intend to use more conducting materials such as iron and steel, makes the condition favorable for lightning occurrences. This study by Lal and Pawar (2011) explained that industrialization increases aerosol concentration in a city which in turn acts as a catalyst in cloud formation and affects cloud's inner workings. Another study by Andrea et al. 2008 confirmed that aerosol increases the convection in cloud up to 50% and peak updraft up to 20% which confirms that lightning will increase with increase in aerosols. Another probable reason for the focus on built land could be the population density. The more the population, the more the heat, more pollutants, and warmer atmosphere that has higher moisture which may be one of the factors for lightning to occur. This may also be a possible reason for the observed concentration of lightning strikes in built areas.

One day for each month with the highest frequency of lightning strikes was selected and compared side by side to the average wind speed and direction of that day and the average temperature (Figure S1). From figure 3 (October 2022), it was observed that, regions where maximum lightning strikes were seen, the wind speed was low around 1-4 m/s probably causing the accumulation of water vapor thus the formation of clouds. It was noted that wind speed decreases when it hits the land area, and more strikes have been observed where wind speed is neither very high nor very low (1-4 m/s). Another pattern observed was that when the wind direction is curving up due to elevation of land, more lightning strikes occur. The speed of winds observed when winds curl, is mostly low (0-2 m/s).

Temperature gradient is also one of the responsible factors in the creation of clouds. Due to warm temperature, hot air rises which causes water droplets to form and with a drop in temperature, condensation of water vapor and subsequent formation of clouds happens. As per figure 4 (June 2022), almost all the observed instances with lightning strikes show that the surface temperature is in 21-30 ˚C range which may provide a conducive environment for lightning occurrences (Price et al. 2023). The comparative plots of lightning strikes of all the months with temperature can be referred to from figure S2.

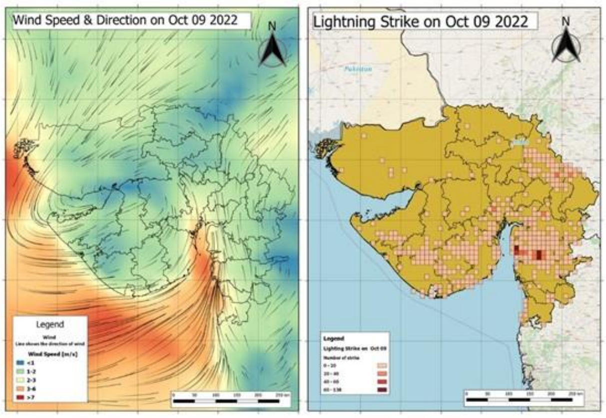


Figure 3. Average Wind map on the day of maximum lightning strikes during October month

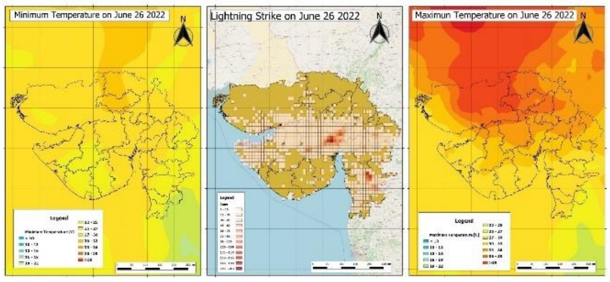


Figure 4. Maps showing temperature gradient at the time of maximum and minimum temperature on the day of maximum lightning strikes in the month of June 2022

Mapping lightning incidents with simulated weather data and finding correlation

A Kernel Density Estimation (KDE) plot as shown in figure 5, maps the frequency of the numerical variables against the numbers repeated (density) in the provided column in the dataset for a particular parameter and draws a continuous and smooth estimation curve (KDE). The figure consists of the kernel distribution plots of different numerical parameters available in the dataset like temperature_2m_mean (average temperature 2m above sea level during the day), frequency of shortwave radiation penetrating the atmosphere and precipitation. The orange curve represents the distribution where the lightning strike did happen, and the blue curve represents the distribution where the lightning strikes did not happen.

One can see from figure 5, that the lightning strikes were more likely to happen at instances where the temperature 2 meters above sea level was in the range 25-30 °C whereas in the same range frequency of no lightning strikes is very low. Lightning strikes are also more likely to happen when the shortwave radiation sum is between 0-

15 W/m^2. Lightning strikes do not occur if the duration of precipitation is less than 2.5 hours.

When the eta-0 evaporation (evapotranspiration from a reference stand of actively growing, well-watered grass, 120mm in height) quantities are in the range 0 – 4 mm/s lightning strikes again are more likely to happen. Wind speed above 2.5-10 m/s was seen to be related to more lightning strikes. No significant dependence of the wind direction was observed on the lightning strikes as the distributions for classes 0 and 1 overlap with each other (Figure S3).

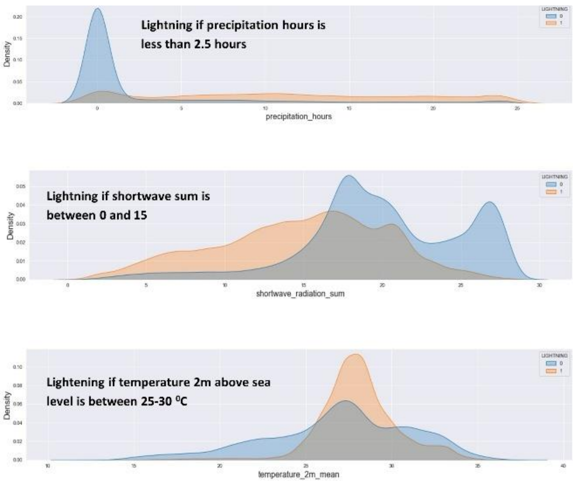


Figure 5. KDE Estimation plot for different parameters

Prediction Model on lightning incidents and comparison of results with real lightning data

We further discuss the results of our prediction model and compare the results against the observations. The model is based on several parameters like temperature, rainfall, shortwave radiation, wind speed, wind direction, snowfall, precipitation, and evapotranspiration. The classification report was generated using a machine learning classification model known as XGBoost, which is a tree-based algorithm (Géron 2019) that uses the attributes provided in the dataset to classify an instance into either a record with lightning strike or a record without lightning strike without excessive scaling.

Precision computes the number of instances correctly classified over the total number of instances, similarly, recall computes the number of correctly classified instances over the total number of instances predicted as the current label. The definition of it is given as

F1-score = 2*precision*recall/ (precision + recall)

More details can be found in Supporting Information.
After the model was trained, the metrics as shown in Figure 6 were obtained.

	precision	recall	f1-score	support
0	0.87	0.86	0.87	11826
1	0.86	0.87	0.86	11352
accuracy			0.86	23178
macro avg	0.86	0.87	0.86	23178
weighted avg	0.87	0.86	0.86	23178

Figure 6. Classification Report

As described before here 0 stands for no lightning strike and 1 stands for a lightning strike. The model is under sampled which means that extra 0 records were eliminated, so the overfitting is not a problem. In this case, the precision of the model predicting 0 is 0.87 which means that 87% of the time the model will predict 0 correctly given that the dataset follows the considered trend. The precision of 1 is also 86 % indicating that 86 percent of the time the model can predict a strike correctly for the given dataset. The recall score tells you what proportion of the actual values it is predicting correctly. So, in this case the recall score of 0 is 86% indicating that out all the 0's 86 percent of them have been classified correctly same goes for the 1's.

The dataset itself includes data points from throughout the year with around 630k records, out of which only 33k are records with a cloud to ground lightning strikes. So almost 500 k zero records were deleted for this model to be developed without overfitting on the dataset. Anomaly detection algorithms were deployed but they unfortunately didn't work as well as we assumed they would, which is an indication that the variation factor that is majorly contributing towards cloud to ground lightning strikes is missing from this dataset.

However, to visualize the separation between data records that had a cloud to ground lightning strike versus data records that don't have strikes, we used principal component analysis (PCA) to plot them on a 3-D graph and see them on our console. Here is a still from the interactive rendered template (Figure 7).

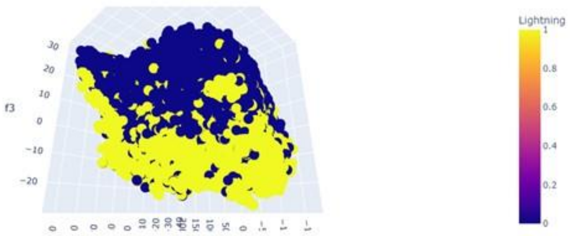


Figure 7. PCA based 3-dimensional visualization of data records with (yellow) and without (blue) lightning strikes.

The yellow points are records where the lightning strikes did happen, and the blue ones are records where they didn't. One can see an approximate plane of separation between the two classes which is what the model is trying to capture as well.

Figure 8 represents a confusion matrix obtained after the model was trained on the training set and then tested on the testing set. A confusion matrix is an efficient way to interpret the accuracy of a classification model and where or how it needs to be modified.

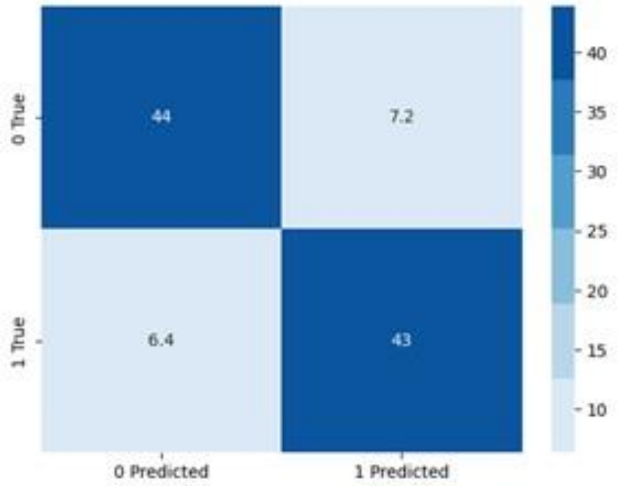


Figure 8. Confusion Matrix

On the y-axis plotted are the actual bins of the lightning strikes and on the x – axis the values predicted by the model. The matrix implies correctly predicted instances across the diagonal.

The model successfully predicted 87 % of the records correctly. There is a 17.1 % chance, however, that the predictions will detect severe lightning, but it will be a false alarm, and a 12.98 % chance that it will be predicted to be low but there will be severe lightning strikes. The probability in the latter half is known as a Type 2 error and such a scenario must be absolutely avoided. 5.46 % in case of life and death is still a severe probability and this number must be reduced.

The model poses the following limitations. (a) The dataset engineered for the model is accumulated and linked through multiple sources. Since the data was unavailable, measures were made to collect data through application programming interfaces (API's) and join them together. (b)Furthermore, the API from which the dataset was generated produces simulated weather data and not actual measurement. The predictions are entirely relied upon the fact that this simulated data is accurate. (c)The time duration of the data used to make the model is from January to December 2022. A model based on such a short time span cannot be relied upon for long-term predictions. (d)When the model was used to perform without deleting records it was not able to distinguish the records as perfectly as it did with the training validation and testing sets. This indicates that the lightning strikes problem could be a time series problem, an approach that we are working on.

4. Summary

The present investigation explores the relationship among the climatic and geographic factors which affect the frequency and distribution of lightning strikes across the state of Gujarat. We note that by incorporating the elevation, rainfall, surface temperature, and winds into the deep learning algorithm, the lightning occurrences over Gujarat can be predicted with more than 80% efficiency. These findings can be further improved by incorporating the real-time weather data input and multi-year database into the model. Thus, developed model can be expanded for nationwide practices which would be helpful to various agencies for carrying out disaster mitigation efforts.

Acknowledgements

Authors acknowledge National Remote Sensing Center (NRSC), Indian Space Research Organization (ISRO), Hyderabad for providing entire lightning data. Authors acknowledge Aman Gupta, Jamin Purohit, Jiya Patel, Nupur Kale, Vedant Mistry, and Aayushi Raval for plotting the lightning data and discussions. We acknowledge deeply Navrachana University for financial support through 'KHOJ Summer Grant 2023', and motivation to conduct this research study.

References

Asfur M., C. Price, J. Silverman and A. Wishkerman (2020) Why is lightning more intense over the oceans? Journal of Atmospheric and Solar-Terrestrial Physics, 202, 105259.

Andreae O.M. and D. Rosenfield (2008) Aerosol-cloud-precipitation interactions. Part 1. The nature and sources of cloud-active aerosols, Earth-Science Reviews, 89(1-2), 13-41.

ArcGIS Homepage, http://www. livingatlas.arcgis.com

Bhuvan, Indian Geo-platform of ISRO, National Remote Sensing Center, NRSC, https://bhuvan-app1.nrsc.gov.in/lightning/

Chen M., X. Gou, Y. Du (2013) The effect of ground altitude on lightning striking distance based on a bi-directional leader model. Atmospheric Research, 125-126, 76-83.

Cecil D.J., D. E. Buechler and R. J. Blakeslee (2014) Gridded lightning climatology from TRMM-LIS and OTD: Dataset description, Atmospheric Research 135–136, 404–414.

Copernicus Climate Data Store Homepage, http://www.cds.climate.copernicus.eu

Cummins K. L., M. J. Murphy, E. A. Bardo, W. L. Hiscox, R. B. Pyle and A. E. Pifer (1998) A combined TOA/MDF technology upgrade of the US National Lightning Detection Network. Journal of Geophysical Research: Atmospheres, 103(D8), 9035-9044.

Dewan A., K. A. Islam, T. R. Fariha, M. M. Murshed, A. Ishtiaque, M. S. G. Adnan and M. B. H. Chowdhury (2022) Spatial pattern and land surface features associated with cloud-to-ground lightning in Bangladesh: an exploratory study. Earth systems and environment, 6(2), 437-451

DIVA-GIS Homepage, http://www. Diva-Gis.org

Géron A. (2019) Machine Learning Algorithms studied from Oreilly's Hands on machine Learning with Sci-kit Learn. 2nd Edn. O'Reilly Media, Inc

Ghosh R., S. D. Pawar, A. Hazra, J. Wilkinson, D. Mudiar, A. Manoj, K. Domkawale, G. Vani and V. Gopalakrishnan (2023) Seasonal Lightning Distributions over Indian Subcontinent. Earth and Space Science, 10(6)

Hazmi A., M. Hamid, R. Fernandez, H. Andre, R. Pratama and P. Emeraldi (2023) The correlation between lightning and various weather parameters in the Padang monsoon system. Indonesian Journal of Electrical Engineering and Computer Science, 31 (1), 1-9

Holle R. L. (2014) Diurnal Variations of NLDN-Reported Cloud-to-Ground Lightning in the United States, Monthly Weather Review, 1037-1052

Holle R. L., W. A. Brooks and K. L. Cummin (2016) Seasonal, Monthly, and Weekly Distributions of NLDN and GLD360 Cloud-to-Ground Lightning, Monthly Weather Review, 144(8), 2855–2870.

Holzworth R. H., M. P. McCarthy, J. B. Brundell, A. R. Jacobson and C. J. Rodger (2019) Global Distribution of Superbolts. Journal of Geophysical Research: Atmospheres, 124, 9996–10,005

Kotroni V. and K. Lagouvardos (2008) Lightning occurrence in relation to elevation, terrain slope, and vegetation cover in the Mediterranean. Journal of Geophysical Research, 113, D21118

Kundu A., S. S. Kundu, S. K. Sharma, M. Gogoi, T. Banik, A. Borgohain, R. Mahanta and A. Debnath (2023) The behavior of cloud base height over a hilly remote station of North-East India using ground-based remote sensing technique. Atmospheric Research, 282, 106512.

Lal D.M. and S. D. Pawar (2011) Effect of urbanization on lightning over four metropolitan cities of India, Atmospheric Environment, 45(1), 191-196

LEE L.S. (2011) Does lightning always strike the tallest object? Hong Kong Observatory web. https://www.hko.gov.hk/en/education/weather/thunderstorm-and-lightning/00013-does-lightning-always-strike-the-tallest-object.html. Accessed on 22 September 2024.

NOAA National Severe Storms Laboratory (http://www.nssl.noaa.gov)

Open-Meteo Homepage, https://open-meteo.com

Oulkar S., D. Singh, U. Saha and A. K. Kamra (2019) Distribution of lightning in relation to topography and vegetation cover over the dry and moist regions in the Himalayas. Journal of Earth System Science, 128, 180(1-17)

Perez-Invernon F.J., F. J. Gordillo-Vazquez, H. Huntrieser and P. Jockel (2023) Variation of lightning-ignited wildfire patterns under climate change. Nature Communications 14, 739

Price C. and D. Rind (1992) A Simple Lightning Parameterization for Calculating Global Lightning Distributions. Journal of Geophysical Research,97(D9), 9919-9933

Raheem B., E. Ogbuju and F. Oladipo (2023) Development of a Lightning Prediction Model Using Machine Learning Algorithm: Survey. Journal of Applied Artificial Intelligence, 4(1), 45-56

Rakov V.A. and M. A. Uman (2003) Lightning: Physics and Effects, Cambridge University Press, Cambridge, U.K. 687 pp, ISBN 0-521-58327-6

Ranalkar M. (2009) Chaudhari H.S.: Seasonal variation of lightning activity over the Indian subcontinent. Meteorology and Atmospheric Physics, 104, 125–134

Rocque M. N., W. Deierling W., K. L. Rasmussen, R. I. Albrecht and B. Medina (2024). Lightning characteristics associated with storm modes observed during RELAMPAGO. Journal of Geophysical Research: Atmospheres, 129,e2023JD039520. https://doi.org/10.1029/2023JD039520

Romps D. M., J. T. Seeley, D. Vollaro and J. Molinari (2014) Projected increase in lightning strikes in the United States due to global warming. Science 346(6211):851–854. https:// doi. org/ 10. 1126/ science. 1259100.

Sharma A., A. Mitra, V. Vasan and R. Govindarajan (2020) Spatio-temporal relationships between rainfall and convective clouds during Indian monsoon through a discrete lens. International Journal of Climatology, 41(2), 1351-1368

Sharma S., A. Dass, A. K. Mishra, S. Singh and K. Kumar (2022) A decadal climatology of cloud vertical structure over the Indo-Gangetic Plain using radiosonde and radar observations. Atmospheric Research, 266, 105949

Singh D., P. S. Buchunde, R. P. Singh, A. Nath, S. Kumar, R. N. Ghodpage (2014) Lightning and convective rain study in different parts of India. Atmospheric Research, 137, 35-48

Singh U., M. Rizwan, M. Alaraj and I. Alsaidan (2021) A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments. *Energies*, 14(16), 5196

Taori A., A. Suryavanshi, B. Gharai and M. V. R. Seshasai (2021) Atmospheric Lightning as Essential Climate Variable (ECV) and Its Availability Over India Using NRSC/ISRO Lightning Detection Sensor Network. Preprints, 2021120183. doi:10.20944/preprints202112.0183.v1.

Taori A., A. Suryavanshi, S. Pawar and M. V. R. Seshasai (2022) Establishment of lightning detection sensors network in India: generation of essential climate variable and characterization of cloud-to-ground lightning occurrences. Natural Hazards 111, 19-32(2022).

Taori A., A. Suryavanshi, R. V. Bothale (2023) Cloud-to-ground lightning occurrences over India: seasonal and diurnal characteristics deduced with ground-based lightning detection sensor network (LDSN). Natural Hazards 116, 4037–4049.

Taori A., A. Suryavanshi, R. Goenka, D. Venkatesh and G. S. Rao (2024) Inter-comparison of Worldwide

Lightning Location Network (WWLLN) and Lightning Detection Sensor Network (LDSN) data over India. Journal of Atmospheric and Solar-Terrestrial Physics, 261, 106286

Shi T., Y. Yang, L. Liu, Y. Tian, Z. Zheng, Y. Huang, Z. Xiao, Y. Wang, Y. Wang and G. Lu (2023) Spatiotemporal patterns of long series of cloud-to-ground lightning in Beijing and its cause. Urban Climate, 49,101480 (2023).

Shi T., G. Lu, X. Wen, L. Liu and P. Qi (2024) A Systematic Review of the Potential Influence of Urbanization on the Regional Thunderstorm Process and Lightning Activity, Atmosphere, 15(3), 374 (2024).

Wang X., K. Hu, Y. Wu and W. A. Zhou (2023) Survey of Deep Learning-Based Lightning Prediction. Atmosphere, *14*, 1698

Williams E.R (1992) The Schumann Resonance: A Global Tropical Thermometer. Science, 256(5060), 1184-1187

Wu M., Z. Sun and Q. Tang (2020) The Variation of Lightning Current Parameters with Altitude in Shandong Region. IOP Conf. Series: Earth and Environmental Science 598, 012084

Yair Y. (2018) Lightning hazards to human societies in a changing climate, Environmental Research Letters 13, 123002

XGBoost Basics, xgboost.com

KDE Estimation, https://jakevdp.github.io/PythonDataScienceHandbook/05.13-kernel-density-estimation.html

PCA Basics, https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html