

# EpiClim: Weekly District-Wise all-India multi-epidemics Climate-Health Dataset for accelerated GeoHealth research

Gurleen Kaur<sup>a</sup>, Shubham Ghoshal<sup>b</sup>, Manmeet Singh<sup>c</sup>, Reena Marbate<sup>d</sup>, Neetiraj Malviya<sup>e</sup>, Arshmehar Kaur<sup>f</sup> and Vaisakh SB<sup>g</sup>

<sup>a</sup>Indian Institute of Science Education and Research, Bhopal, India ,  
Banaras Hindu University ,

<sup>c</sup> Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA,

<sup>d</sup>Indian Institute of Public Health, Delhi, India ,

<sup>e</sup>Centre for Development of Advanced Computing, Pune, India ,

<sup>f</sup>Guru Teg Bahadur Institute of Technology, Delhi, India ,

<sup>g</sup>Indian Institute of Tropical Meteorology, Pune, India ,

arXiv:2501.18602v1 [physics.soc-ph] 17 Jan 2025

## ARTICLE INFO

### Keywords:

Climate-induced epidemics, climate-health models, epidemic forecasting, disease surveillance, AI-Numerical Weather Prediction (AI-NWP), public health and climate science, vector-borne diseases, district-wise health data, climate change and disease spread, predictive modeling, health datasets for climate research, dengue and malaria modeling, epidemic-climate dataset, coupled climate-health systems, India public health data.

## ABSTRACT

Climate change significantly impacts public health, serving as a critical precursor to the emergence and spread of epidemics. The development of climate health models has become imperative for accurately assessing and predicting climate-induced disease outbreaks. Historical and contemporary evidence underscores the influence of climatic variables, such as temperature and precipitation, on the prevalence of vector- and water-borne diseases. For example, dengue and malaria are strongly correlated with temperature changes, while cholera outbreaks are often linked to precipitation anomalies. Advancements in weather and climate sciences, particularly through AI-enabled numerical weather prediction (AI-NWP), have increased confidence in forecasting key variables like temperature and precipitation. However, a major obstacle to the integration of climate models with health prediction systems lies in the inaccessibility of comprehensive, publicly accessible health datasets for various diseases. Such data sets, especially on granular spatial and temporal scales, are crucial to advance climate health research and model development. Here, we present the first weekly district-wise dataset for major epidemics across India spanning 2009 to the present. The data utilized in this study has been sourced from the Integrated Disease Surveillance Programme (IDSP) portal, a publicly available database maintained by the Government of India. This dataset, tentatively named "EpiClim: India's Epidemic-Climate Dataset", bridges the gap by providing actionable health data tailored for climate-health modeling. The data set offers insight into the temporal and spatial dynamics of diseases such as dengue, malaria, and acute-diarrheal disease. The availability of such datasets paves the way for integrating climate forecasts with epidemic prediction models, enabling actionable insights for policymakers, public health officials, and researchers. This work marks a foundational step toward coupling predictive climate health models with numerical weather and climate models, driving innovation in understanding and mitigating climate-induced public health crises.

## 1. Introduction

Climate change and its associated impacts on public health have emerged as critical global concerns. The interplay between climatic variables, such as temperature and precipitation, and disease dynamics has intensified the prevalence of epidemics such as malaria, dengue, and acute diarrheal diseases. Despite advances in weather and climate science, particularly in artificial intelligence-enabled numerical weather prediction (AI-NWP), accurate prediction and mitigation of climate-induced epidemics remain a challenge.

The development of predictive climate health models is essential for assessing the onset and spread of climate-induced diseases. Such models require robust health data sets integrated with climate variables. However, the lack of detailed publicly available epidemiological data impedes progress in this domain. To bridge this gap, we present a

\*Correspondence to: Banaras Hindu University

✉ gurleen21@iiserb.ac.in (G. Kaur); 23jibs-sdhingra@jgu.edu.in (S. Ghoshal); manmeet.singh@utexas.edu (M. Singh); reenamarbate14796@gmail.com (R. Marbate); neetirajmalviya@gmail.com (N. Malviya); arshmeharkaur05@gmail.com (A. Kaur); vaisakh.sb@tropmet.res.in (V. SB)

ORCID(s):

district-wise, weekly epidemiological dataset for India spanning from 2009 to the present. This dataset captures the dynamics of multiple diseases, enabling spatial and temporal analyses and facilitating the development of climate-health models. The work aims to:

1. Investigate the role of climatic factors in disease spread.
2. Presenting an analytical report for the developed dataset providing in-depth insights and conclusions for epidemic diseases
3. Provide a comprehensive dataset for public health research.
4. Advance the integration of climate models with health prediction systems to enable informed public health interventions.

## 2. Data and Methodology

The methodology comprises the collection, curation, and visualization of disease outbreak data integrated with climatic variables. The dataset includes 15 features covering epidemiological, geographic, and climatic data. Notable variables include:

- Temporal Data: Week, month, and year of outbreaks.
- Geographic Data: State, district, latitude, and longitude.
- Epidemiological Data: Number of Diseases, cases, and deaths.
- Climatic Data: Precipitation (preci), Leaf Area Index (LAI), and temperature (Temp).

The methodology for this project involved a systematic approach to collecting, processing, and visualizing epidemiological data, integrated with relevant climatic variables. First, the epidemiological data was gathered from publicly available reports and cross-referenced with climate datasets to ensure accuracy and consistency.

The data preprocessing phase focused on converting raw entries into formats suitable for analysis, such as transforming strings to numerical formats where applicable (e.g., converting "Cases" to integers) and handling missing values (e.g., NaNs in the "Deaths" field). The final step involved extensive visualization to extract actionable insights. Using Python, spatial distribution maps for 2022 were generated to identify disease hotspots, while temporal trend analyses were conducted to study diseases like Acute Diarrheal Disease across different years. Furthermore, year-wise subplots provided comparative visualizations for selected diseases, covering years like 2011, 2013, and 2015. These steps collectively aimed to uncover spatial and temporal trends, facilitating the development of predictive climate-health models. Further details on the technical validation and coding will refine this methodology.

## 3. Data Description and Technical Validation

### 3.1. Dataset Overview

The EpiClim dataset was meticulously compiled from open sources, with a significant contribution from the Integrated Disease Surveillance Programme (IDSP) portal<sup>1</sup>. The IDSP provides weekly outbreak reports for various diseases across India. We extracted district-wise health data from these reports, including diseases such as Dengue, Malaria, Cholera, and Acute Diarrhoeal Disease. District names were mapped to their geographic coordinates (latitude and longitude) to facilitate geospatial analysis.

To supplement the health data, we incorporated relevant climatic variables such as temperature, precipitation, and Leaf Area Index (LAI) from publicly available climate datasets. These variables were carefully selected based on their potential influence on disease dynamics, as indicated in existing literature. The integration of health and climatic data enabled the formulation of a robust dataset, suitable for exploring the complex interactions between climate and health.

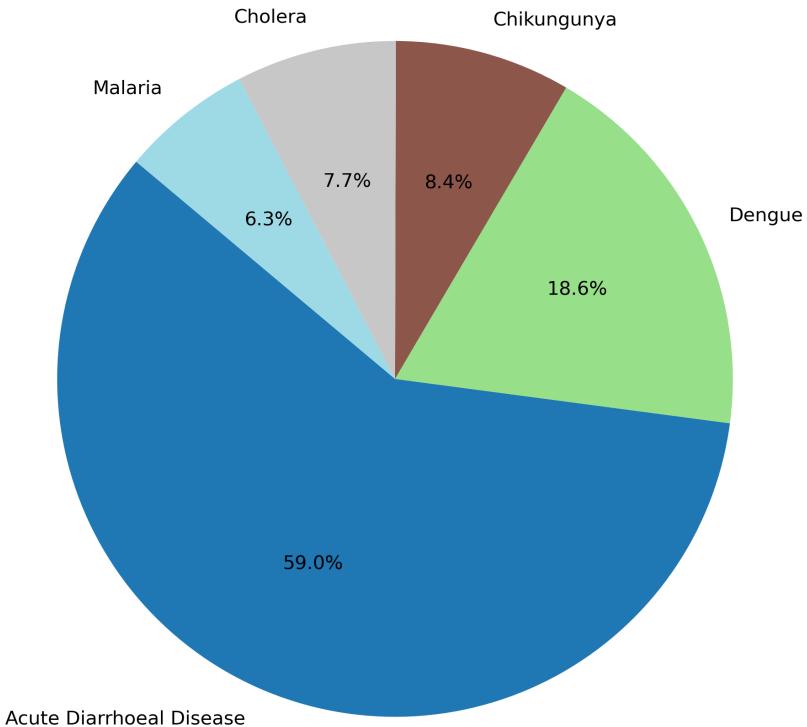
### 3.2. Data Summary

Table 1 summarizes the top diseases and states by cases, along with climatic factor statistics. The most prevalent disease recorded in the dataset is Acute Diarrhoeal Disease, followed by Dengue and Cholera. West Bengal reports the highest number of cases among states. Climatic factors like temperature and precipitation show significant variations, with average temperatures around 304.5K and mean daily precipitation of 0.46mm.

---

<sup>1</sup><https://idsp.mohfw.gov.in/index4.php?lang=1&level=0&linkid=406&lid=3689>

Top 5 Diseases by Occurrence



**Figure 1:** Distribution of top diseases by reported cases. Acute Diarrhoeal Disease accounts for the largest proportion.

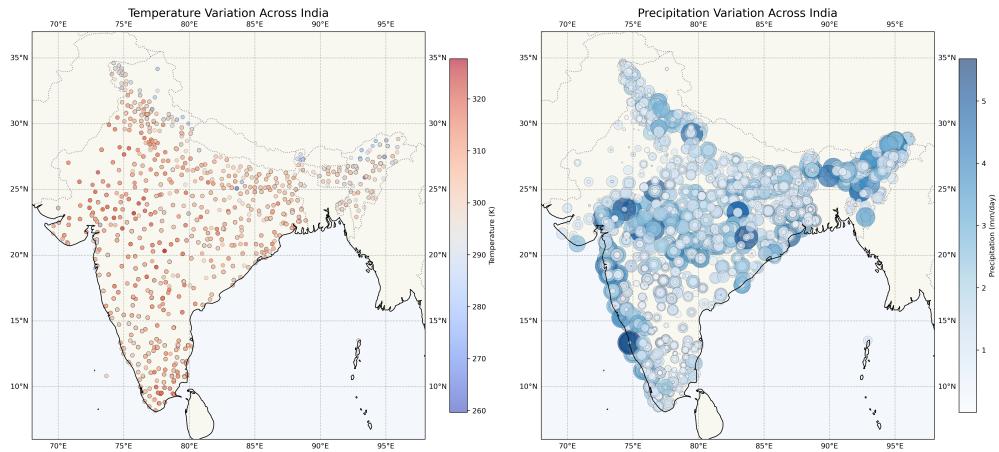
### 3.3. Visualization and Analysis

The dataset is enriched with key visualizations that highlight its scope and applicability. The codebase for the visualisation can be found at [GitHub Repository](#) :

1. **Temporal Trends of Cases and Deaths:** Figures ?? and 3 illustrate the temporal dynamics of reported cases and deaths, smoothed over a 100-day moving average. Both metrics demonstrate periodic peaks, indicative of seasonality in epidemic occurrences.
2. **Disease Prevalence Distribution:** The proportional distribution of top diseases is depicted in Figure 1. Acute diarrheal disease dominates the data set, comprising 59% of the total cases.

### 3.4. Technical Validation

To validate the dataset, we conducted statistical checks for completeness, consistency, and range integrity across all fields. The climatic variables—temperature, precipitation, and LAI—were cross-referenced against publicly available datasets for accuracy. Disease case counts were verified with aggregated district and state-level health reports. The dataset is suitable for climate-health modeling due to its granularity, temporal span, and reliability. The dataset highlights seasonal and regional variations in disease prevalence, strongly influenced by climatic factors. The observed correlations between temperature, precipitation, and epidemic cases suggest significant opportunities for predictive modeling. This dataset serves as a cornerstone for integrating climate forecasts with epidemic prediction models, paving the way for proactive public health interventions.



**Figure 2:** Geospatial variations in temperature (top) and precipitation (bottom) across India.

**Table 1**

Excerpts from the EpiClim Dataset

Disease	Cases
Acute Diarrhoeal Disease	251,456
Dengue	238,047
Cholera	126,495
Malaria	111,858
Chikungunya	53,289

State/UT	Cases
West Bengal	178,830
Delhi	80,933
Uttar Pradesh	60,173
Maharashtra	53,575
Karnataka	39,532

## 4. Results and Discussion

### 4.1. Temporal Trends of Disease Cases and Deaths

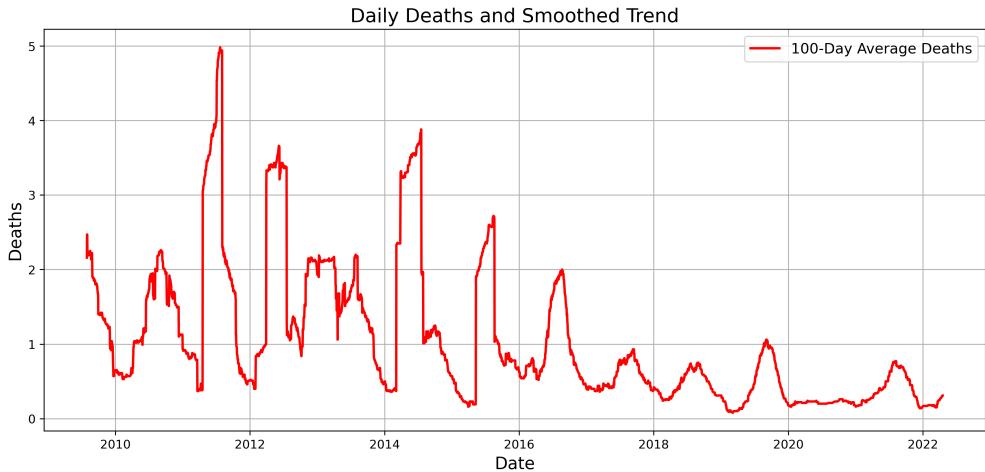
Figure 4 illustrates the temporal trends in yearly deaths across all diseases. A significant decline is observed from 2010 to 2022, suggesting improved public health interventions and increased awareness. Despite this decline, periodic surges in disease cases highlight the persistent seasonal nature of epidemics, necessitating continued monitoring and timely response.

### 4.2. Spatial Distribution of Epidemic Diseases

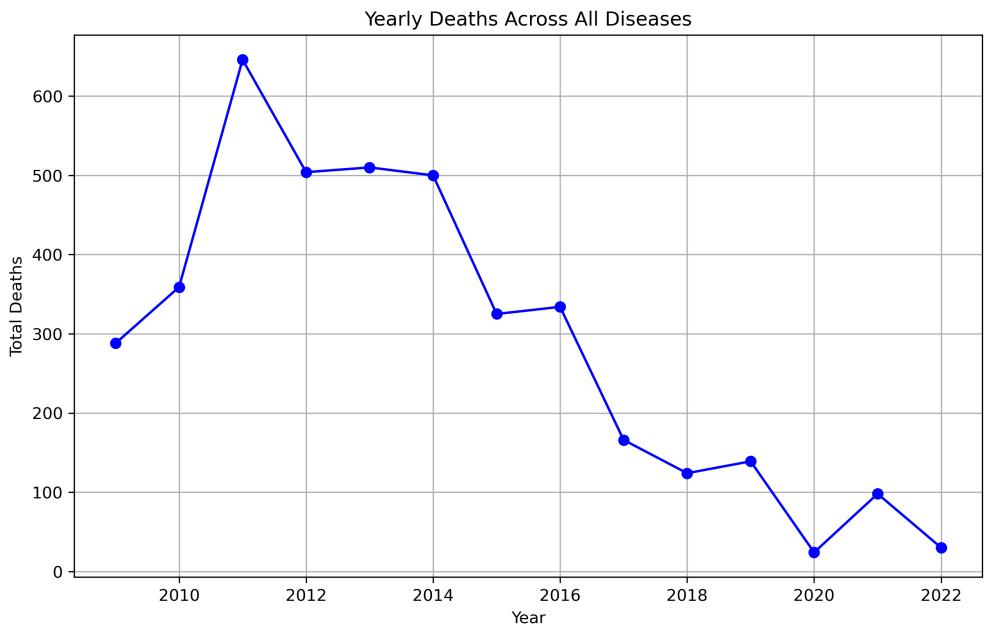
The spatial distribution of Acute Diarrhoeal Disease cases in India (Figure 5) highlights a high concentration of cases in northern and eastern states. This is consistent with regions that experience heavy monsoons and poor sanitation infrastructure, reinforcing the link between waterborne diseases and environmental conditions.

### 4.3. Disease-Specific Trends

The temporal and spatial patterns of specific diseases are depicted in Figures 8, 10, 12, and 14. For Acute Diarrhoeal Disease (ADD), Figure 8 illustrates a steady increase in reported cases up until 2019, followed by a decline in recent years. This trend may signify the success of public health campaigns, although certain high-risk areas continue to experience disease hotspots. Chikungunya patterns, as shown in Figure 10, reveal sporadic outbreaks concentrated in the southern and western regions, where favorable conditions for mosquito breeding persist. Cholera trends, presented in Figure 12, exhibit cyclical outbreaks predominantly in coastal states that are prone to heavy rainfall and flooding.



**Figure 3:** Temporal trends of daily reported deaths, smoothed over a 100-day average.



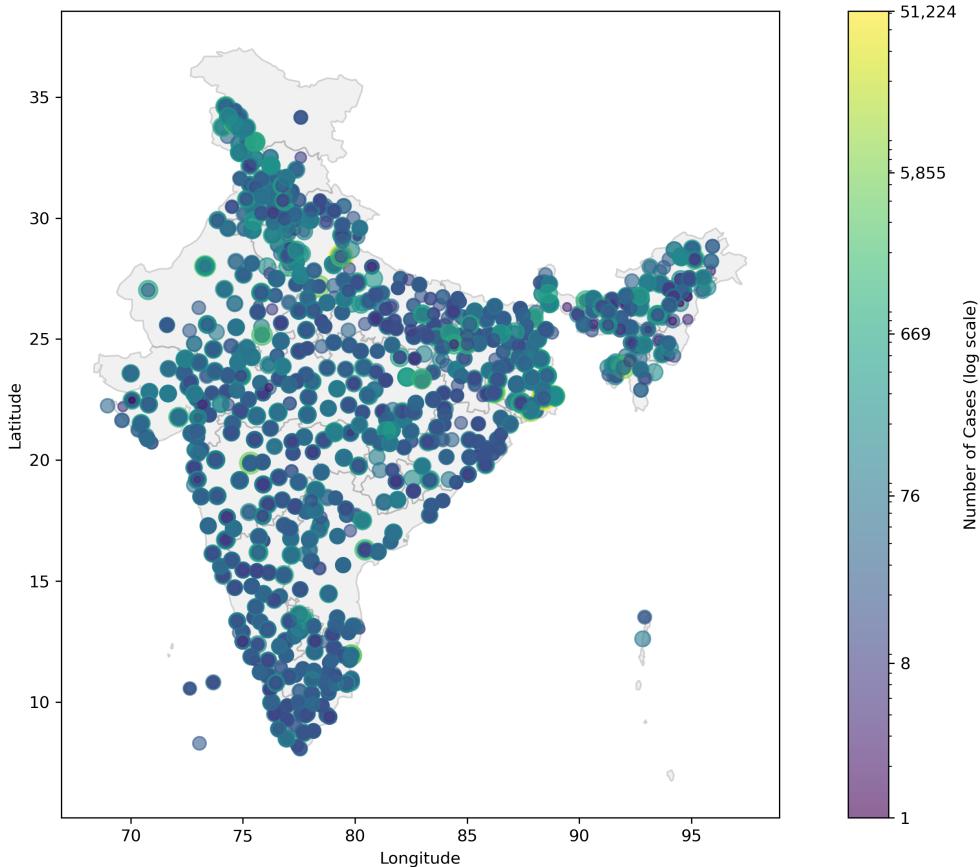
**Figure 4:** Yearly deaths across all diseases show a decreasing trend from 2010 to 2022.

Finally, Fig. 8 highlights the widespread geographic distribution of Figure 14, with cases consistently reported across numerous states. This underscores the expanding habitat range of Aedes mosquitoes, which are vectors for the disease.

#### 4.4. State-Wise Analysis

Figure 6 ranks the top 10 states by total disease cases. West Bengal, Delhi, and Uttar Pradesh report the highest disease burden, consistent with their high population densities and varying climatic conditions. Implementing targeted interventions that incorporate climate driven actions in these states can potentially lead to a substantial reduction in the national disease burden thereby significantly improving public health outcomes.

## Spatial Distribution of Acute Diarrheal Disease Cases in India (2022)



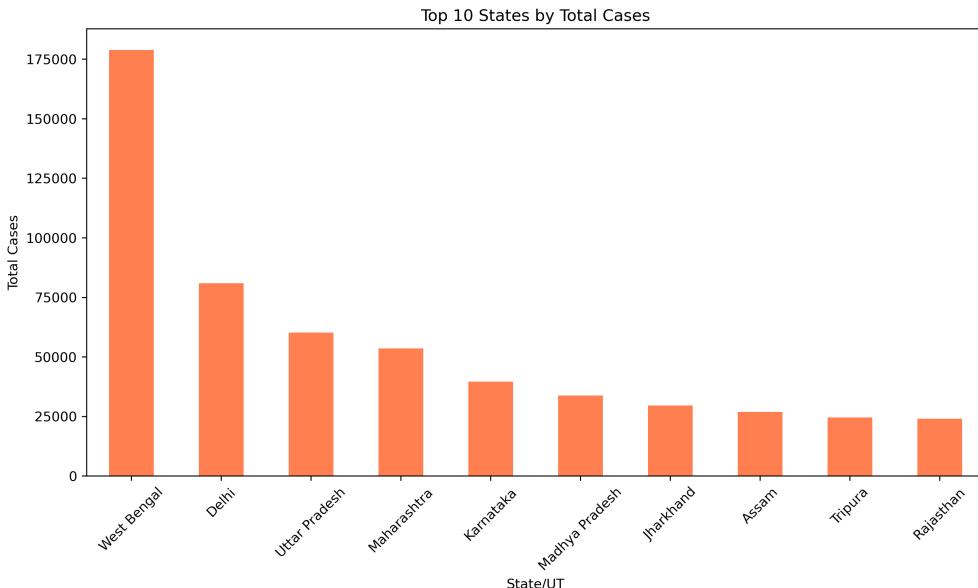
**Figure 5:** Spatial distribution of Acute Diarrhoeal Disease cases in India (2022).

#### 4.5. Discussion and Implications

The results underscore the interplay between climatic factors and disease outbreaks. Spatial and temporal analyses reveal distinct patterns influenced by monsoons, temperature variations, and urbanization. Acute Diarrhoeal Disease and Cholera are heavily influenced by waterborne pathways, while vector-borne diseases such as Dengue and Chikungunya are closely tied to mosquito-breeding conditions.

The decreasing trend in deaths highlights progress in healthcare delivery, yet persistent outbreaks point to gaps in prevention and early detection. This dataset and associated visualizations provide actionable insights for policymakers to design targeted interventions, improve sanitation, and implement climate-resilient public health strategies. Future work could explore machine learning models to predict outbreaks based on climatic variables, enabling proactive mitigation and resource allocation.

- Spatial Hotspots:** Regions with high precipitation levels and dense vegetation, as indicated by the Leaf Area Index (LAI), exhibit a marked increase in the prevalence of vector-borne diseases such as malaria and dengue. These areas act as breeding grounds for disease vectors, underscoring the role of geographic and environmental factors in disease propagation.
- Temporal Patterns:** The data highlights seasonal surges in disease outbreaks, particularly during the monsoon months. This trend aligns with the proliferation of favorable conditions for vector growth, such as increased humidity and standing water. Such patterns demonstrate the seasonality of disease dynamics and the influence of short-term climatic changes.



**Figure 6:** Top 10 states by total disease cases.

**3. Climate-Disease Linkage:** Correlation analyses reveal a strong relationship between climatic variables, particularly temperature and precipitation, and the frequency of disease outbreaks. Higher temperatures, coupled with elevated moisture levels, contribute to vector activity and the subsequent rise in disease incidence.

These findings emphasize the critical need for integrating climate and health data to enhance the prediction and management of epidemic outbreaks. The spatial and temporal visualizations derived from this dataset offer invaluable insights for proactive public health planning, including targeted health interventions, resource allocation, and the development of robust climate-health models. By leveraging these insights, policymakers and researchers can better address the challenges posed by climate-induced epidemics, fostering resilience in public health systems.

## 5. Discussions & Implications

The results underscore the interplay between climatic factors and disease outbreaks. Spatial and temporal analyses reveal distinct patterns influenced by monsoons, temperature variations, and urbanization. Acute Diarrhoeal Disease and Cholera are heavily influenced by waterborne pathways, while vector-borne diseases such as Dengue and Chikungunya are closely tied to mosquito-breeding conditions.

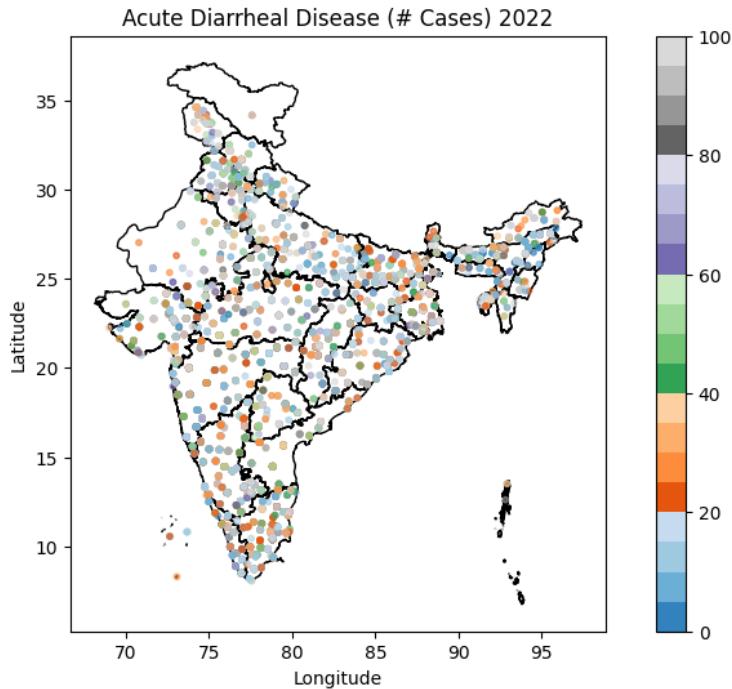
The decreasing trend in deaths highlights progress in healthcare delivery, yet persistent outbreaks point to gaps in prevention and early detection. This dataset and associated visualizations provide actionable insights for policymakers to design targeted interventions, improve sanitation, and implement climate-resilient public health strategies. Future work could explore machine learning models to predict outbreaks based on climatic variables, enabling proactive mitigation and resource allocation.

### 5.1. Spatial Hotspots

Regions with high precipitation levels and dense vegetation, as indicated by the Leaf Area Index (LAI), exhibit a marked increase in the prevalence of vector-borne diseases such as malaria and dengue. These areas act as breeding grounds for disease vectors, underscoring the role of geographic and environmental factors in disease propagation.

### 5.2. Temporal Patterns

The data highlights seasonal surges in disease outbreaks, particularly during the monsoon months. This trend aligns with the proliferation of favorable conditions for vector growth, such as increased humidity and standing water. Such patterns demonstrate the seasonality of disease dynamics and the influence of short-term climatic changes.



**Figure 7:** Acute Diarrheal Case for 2022

### 5.3. Climate-Disease Linkage

Correlation analyses reveal a strong relationship between climatic variables, particularly temperature and precipitation, and the frequency of disease outbreaks. Higher temperatures, coupled with elevated moisture levels, contribute to vector activity and the subsequent rise in disease incidence.

These findings emphasize the critical need for integrating climate and health data to enhance the prediction and management of epidemic outbreaks. The spatial and temporal visualizations derived from this dataset offer invaluable insights for proactive public health planning, including targeted health interventions, resource allocation, and the development of robust climate-health models. By leveraging these insights, policymakers and researchers can better address the challenges posed by climate-induced epidemics, fostering resilience in public health systems.

Figure 17 reveals distinct clustering patterns of disease cases based on temperature and precipitation. Higher case counts are observed within specific ranges of temperature (approximately 300-350 K) and precipitation, suggesting the influence of these environmental factors on disease prevalence.

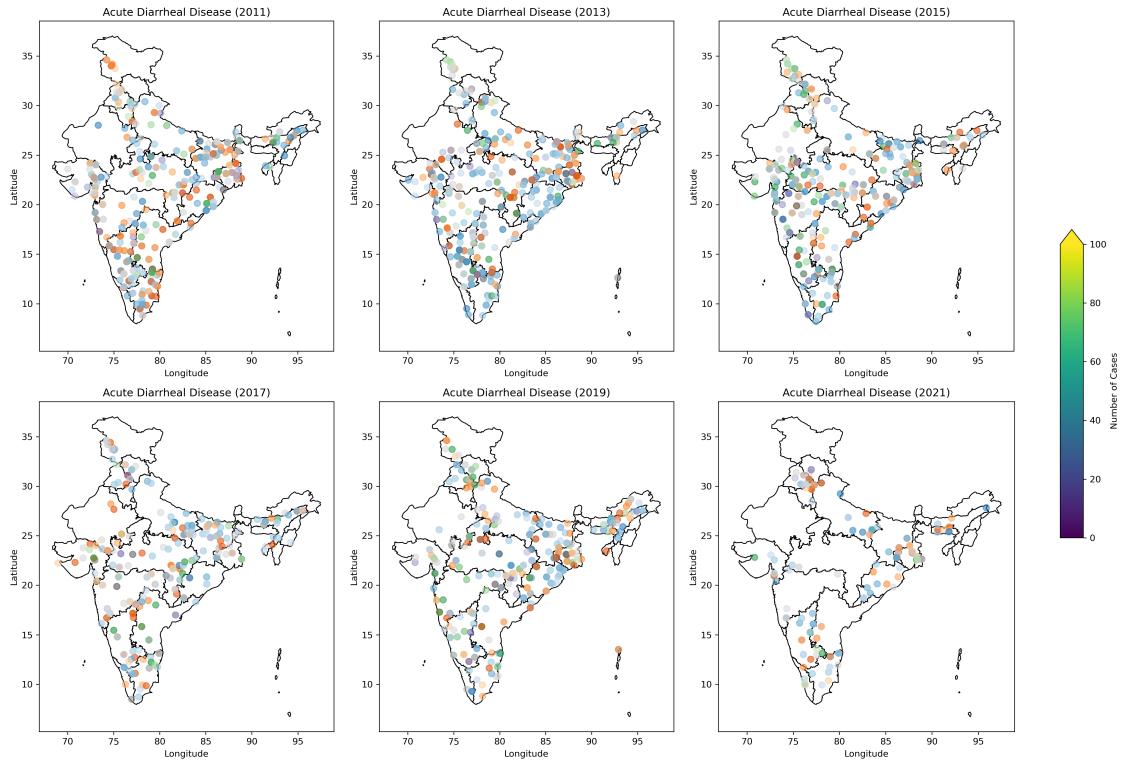
## 6. Results and Analysis

The analysis highlights the temporal and environmental dynamics influencing disease trends. The cumulative monthly trends reveal strong seasonality across four diseases: Acute Diarrhoeal Disease, Cholera, Dengue, and Malaria. Acute Diarrhoeal Disease and Cholera consistently peak during the pre-monsoon and early monsoon periods, suggesting a link to water contamination and inadequate sanitation. In contrast, Dengue and Malaria exhibit spikes during and after the monsoon, aligning with favorable breeding conditions for mosquitoes.

Smoothed monthly averages refine these patterns, offering a clearer depiction of seasonal peaks. Dengue consistently peaks in September, while Malaria shows a broader post-monsoon rise. Both waterborne and vector-borne diseases exhibit distinct dependencies on climatic factors, underscoring the importance of environmental conditions in driving their spread.

Yearly trends, analyzed through cumulative precipitation and disease cases, reveal interannual variability. Precipitation peaks in 2014 and 2018 coincide with higher Malaria cases, confirming a direct relationship. However, Dengue's sharp rise in 2016 does not correspond with precipitation alone, suggesting additional factors like vector

## Climate Health Nexus



**Figure 8:** Temporal trends in Acute Diarrheal Disease cases across multiple years.

control measures or population dynamics may have played a role. Over the years, Dengue shows a rise until 2016 followed by a decline, possibly due to improved public health interventions. Malaria exhibits a steady decline, reflecting successful control efforts. On the other hand, Acute Diarrhoeal Disease and Cholera show more stable trends, indicating persistent issues related to water and sanitation.

In summary, the analysis underscores the strong influence of seasonality and precipitation on disease outbreaks. While some diseases like Malaria show progress in control efforts, others, such as Dengue, demand continued vigilance and targeted interventions to mitigate their impact. These findings highlight the importance of integrating climatic and environmental monitoring into public health strategies.

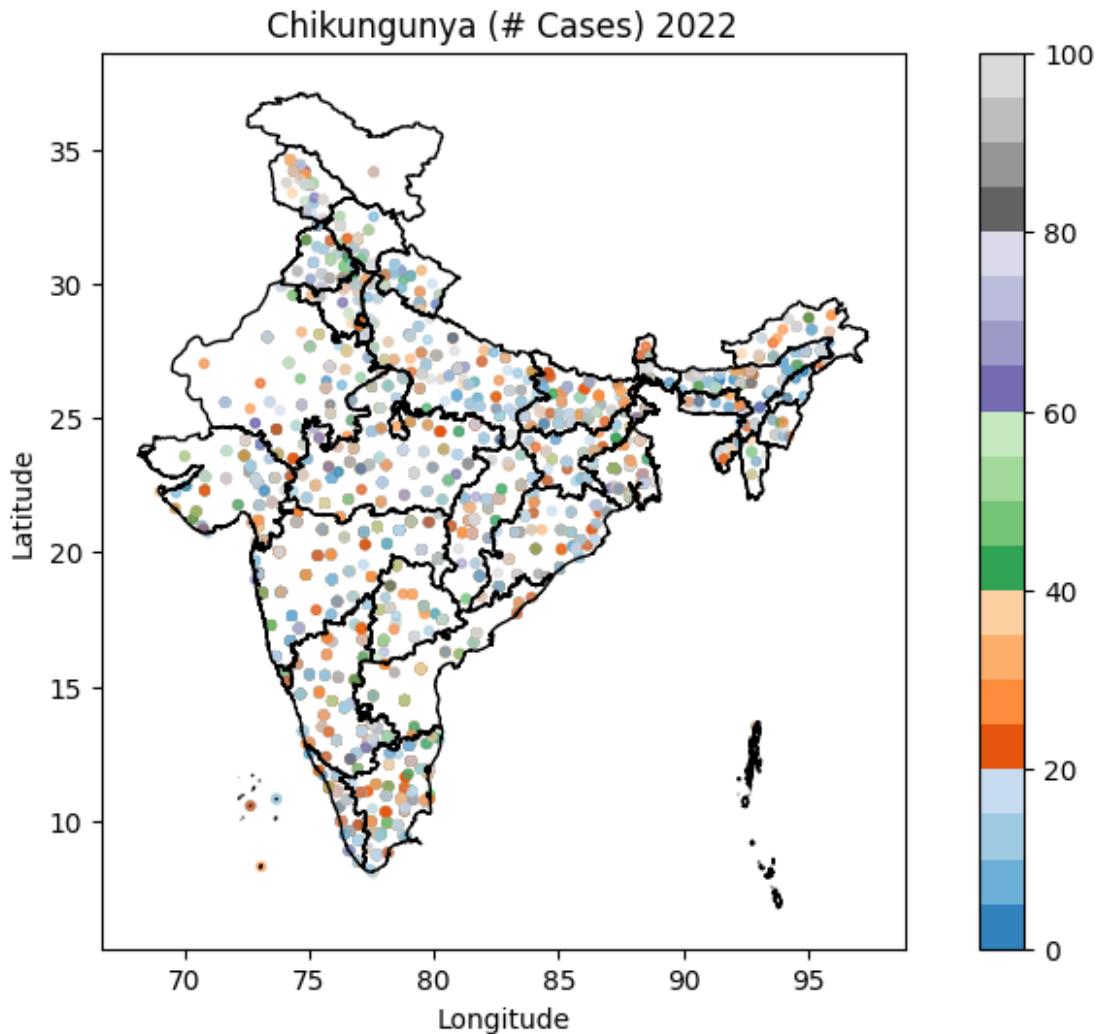
## 6.1. APPENDIX

The figure showcases the spatial distribution and prevalence of four key diseases—Acute Diarrheal Disease (ADD), Chikungunya, Dengue, and Cholera—across India for the year 2022. Each subplot represents one disease, with data visualized as colored dots indicating case counts at specific geographic locations. The legend provides a scale ranging from lower to higher cases, with darker shades denoting higher prevalence.

### 6.1.1. Spatial Patterns and Disease-Specific Observations

**Acute Diarrheal Disease (ADD):** The distribution of ADD cases is widespread across the country, with clusters of high cases in states with dense populations and significant water quality challenges. This pattern highlights the role of sanitation and water access in influencing ADD prevalence, particularly in northern and eastern India. The concentration of cases in states like Uttar Pradesh and Bihar suggests regional disparities in public health infrastructure and waterborne disease control measures.

**Chikungunya:** Chikungunya cases are less uniformly distributed compared to ADD, with higher concentrations observed in southern and western states, such as Karnataka and Maharashtra. This pattern is correlated with the climatic factors that favor the breeding of Aedes mosquitoes, such as warm temperatures and seasonal rainfall. The spatial variation underscores the need for region-specific vector control programs.



**Figure 9:** Spatial distribution of Chikungunya cases in India in 2022.

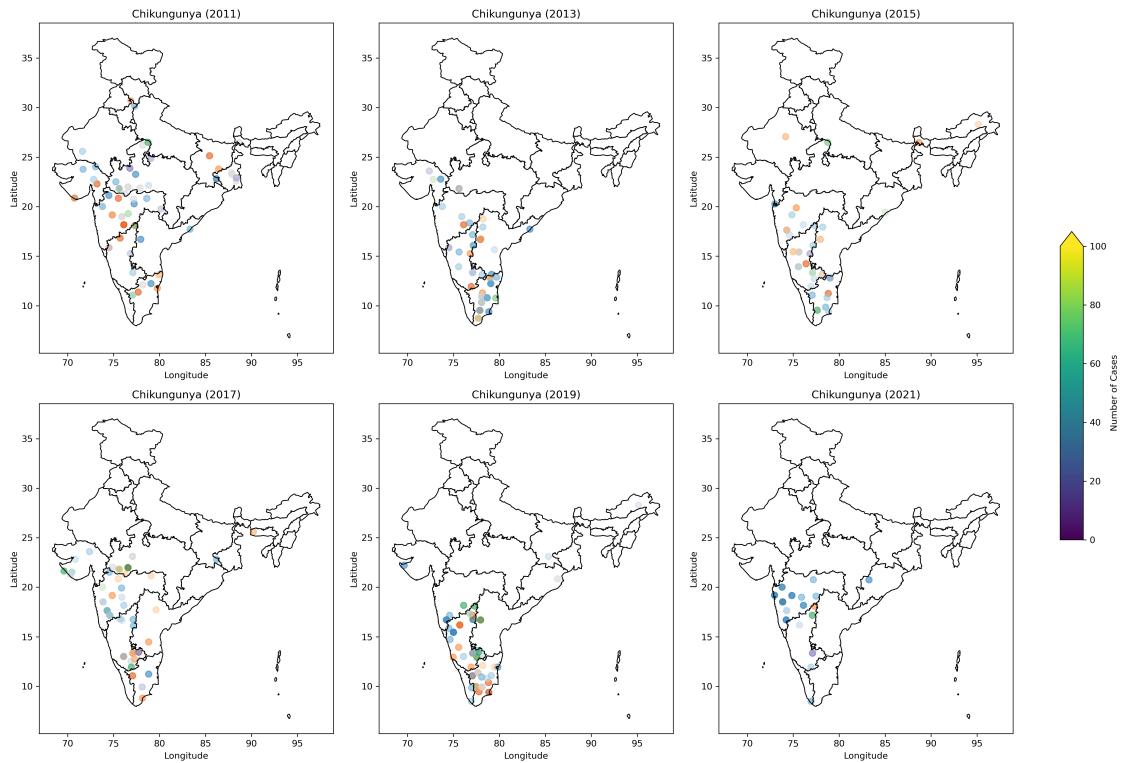
**Dengue:** Dengue cases exhibit a more widespread pattern, with a significant burden across both northern and southern regions of the country. Urban centers like Delhi and Mumbai likely report high cases due to rapid urbanization, poor waste management, and favorable breeding conditions for mosquitoes. This disease's prevalence across diverse climatic zones reflects its multifaceted nature, influenced by urbanization and climate factors.

**Cholera:** Cholera cases are primarily concentrated in eastern and northeastern states, such as West Bengal and Assam, regions often affected by flooding and inadequate sanitation infrastructure. This clustering highlights the intersection of climatic vulnerabilities, such as monsoonal flooding, with public health challenges in these regions.

#### **6.1.2. Regional and Climatic Drivers**

The maps collectively emphasize the influence of climatic and geographic variability on disease prevalence. High-density population areas and regions prone to extreme weather events, such as flooding, emerge as hotspots for multiple diseases. Additionally, regions with warmer climates and seasonal rainfall patterns show a higher burden of vector-borne diseases like Dengue and Chikungunya.

## Climate Health Nexus



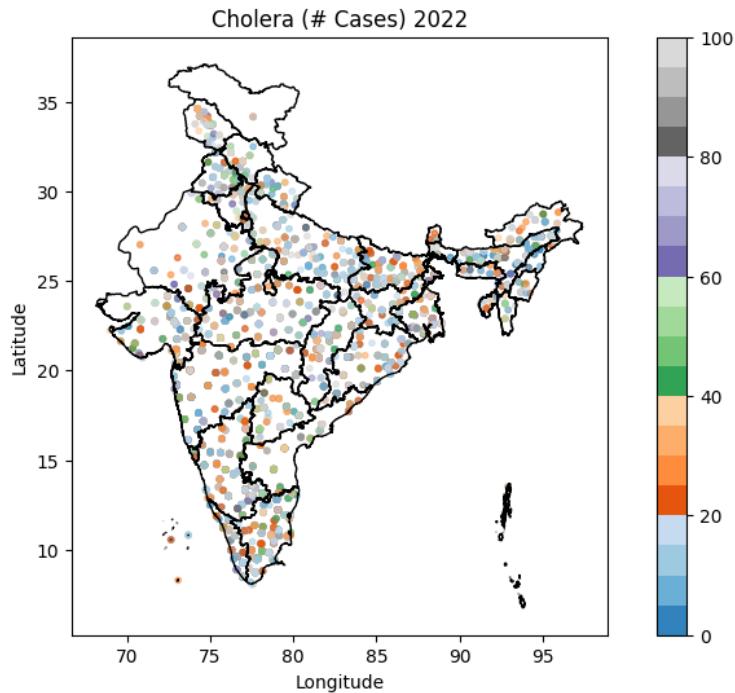
**Figure 10:** Spatial distribution of Chikungunya cases in India across selected years.

### 6.1.3. Implications for Public Health

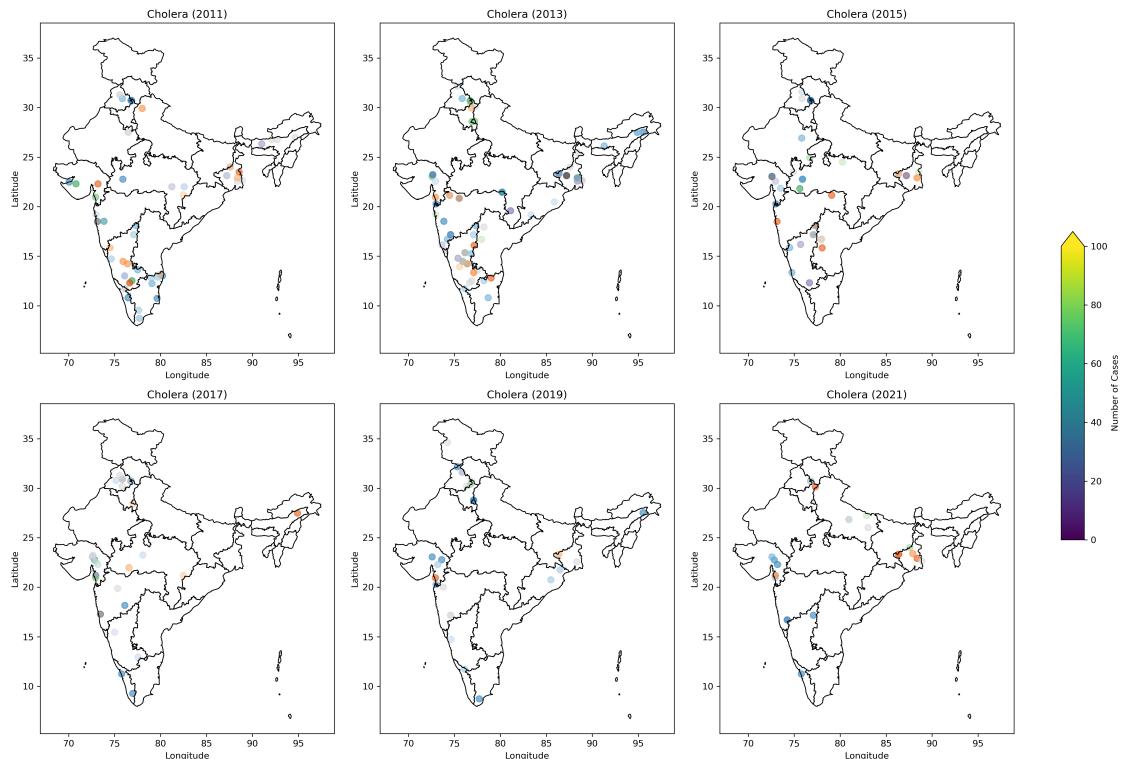
This spatial analysis highlights the critical need for targeted interventions and improved public health infrastructure. States with recurrent high disease burdens, such as West Bengal, Uttar Pradesh, and Delhi, should prioritize integrated disease management strategies, including improved sanitation, robust vector control measures, and early-warning systems tied to climate forecasts. Addressing regional vulnerabilities through climate-sensitive public health planning could significantly reduce the national disease burden and foster resilience against climate-induced health crises.

## REFERENCES

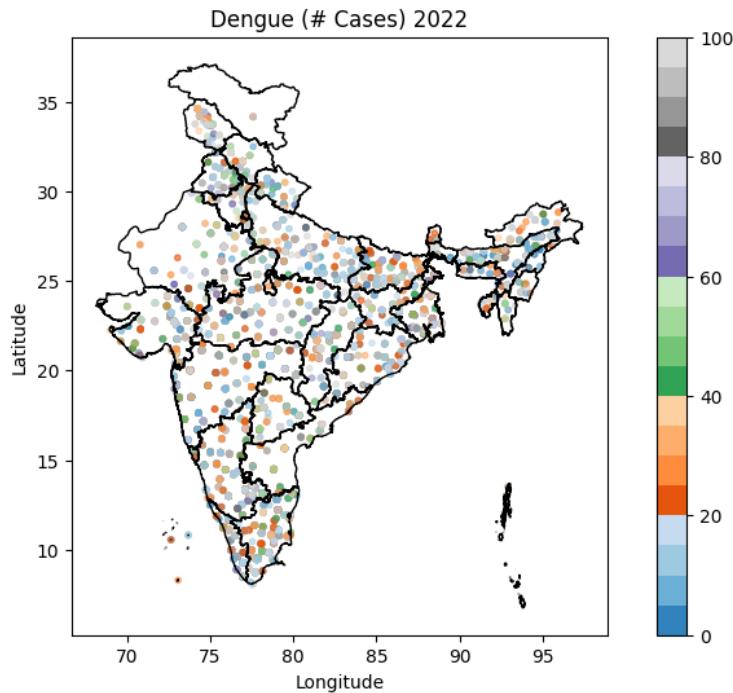
1. Patz, J. A., et al. (2005). "Impact of regional climate change on human health." *Nature*. <https://www.nature.com/articles/nature04188>
2. Levy, K., Woster, A. P., Goldstein, R. S., & Carlton, E. J. (2016). "Untangling the impacts of climate change on waterborne diseases: A systematic review of relationships between diarrheal diseases and climatic factors." *Environmental Science & Technology*.
3. Yang, K., et al. (2021). "Applications of AI in climate modeling and prediction." *Patterns*.
4. <https://idsp.mohfw.gov.in/index.php> This web portal is for online reporting under Integrated Disease Surveillance Programme (IDSP), one of the major National Health Programme under National Health Mission for all States & UTs.
5. Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). "An overview of CMIP5 and the experiment design." *Bulletin of the American Meteorological Society*.
6. India: Health of the Nation's States – The India State-Level Disease Burden Initiative (2017). " Indian Council of Medical Research.
7. Haines, A., et al. (2006). "Climate change and human health: Impacts, vulnerability, and public health." *The Lancet*.
8. Bouma, M. J., & Dye, C. (1997). "Cycles of malaria associated with El Niño in Venezuela." *Journal of the American Medical Association (JAMA)*.



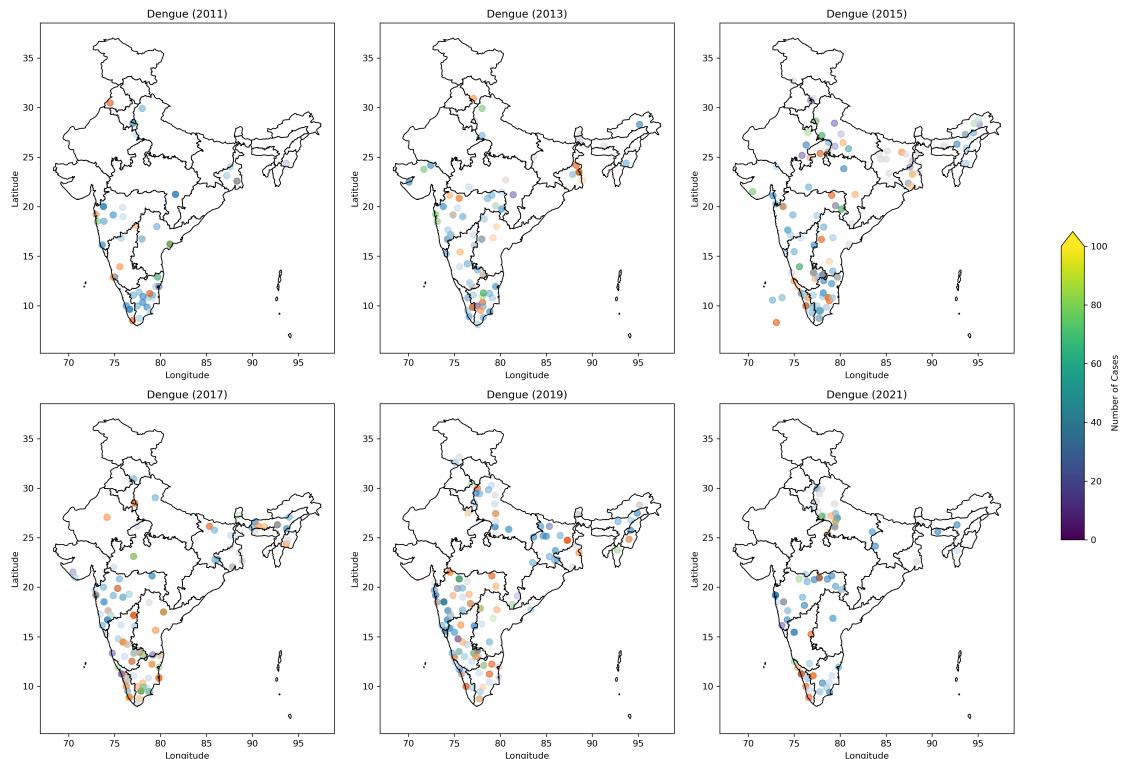
**Figure 11:** Spatial distribution of Cholera cases in India in 2022.



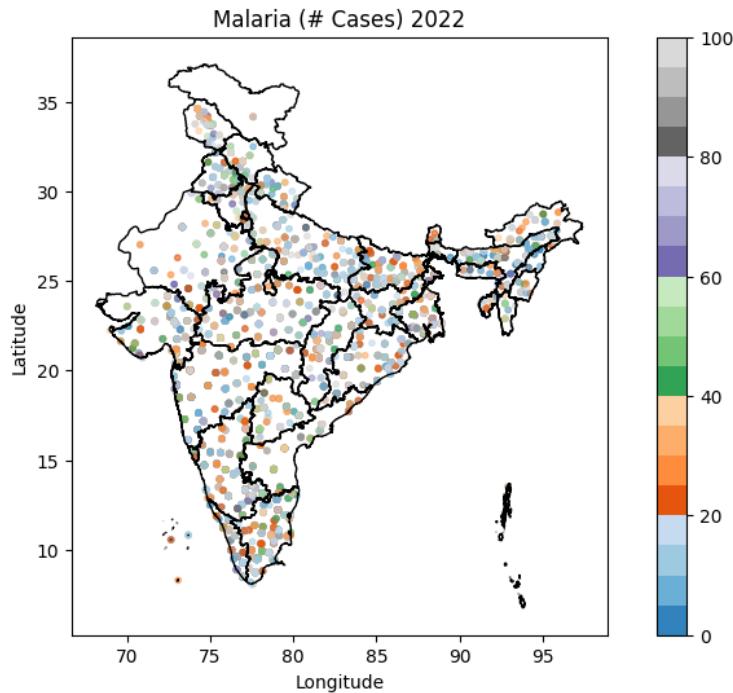
**Figure 12:** Spatial distribution of Cholera cases in India across selected years.



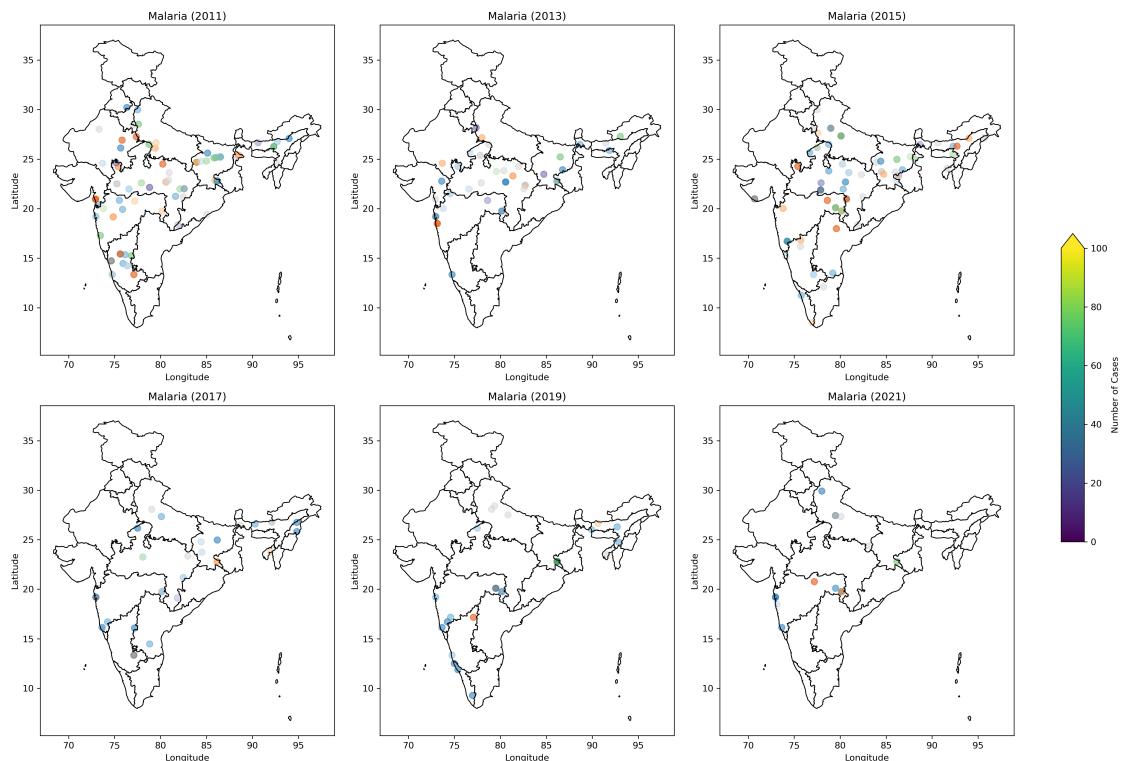
**Figure 13:** Dengue cases for 2022



**Figure 14:** Spatial distribution of Dengue cases in India across selected years.

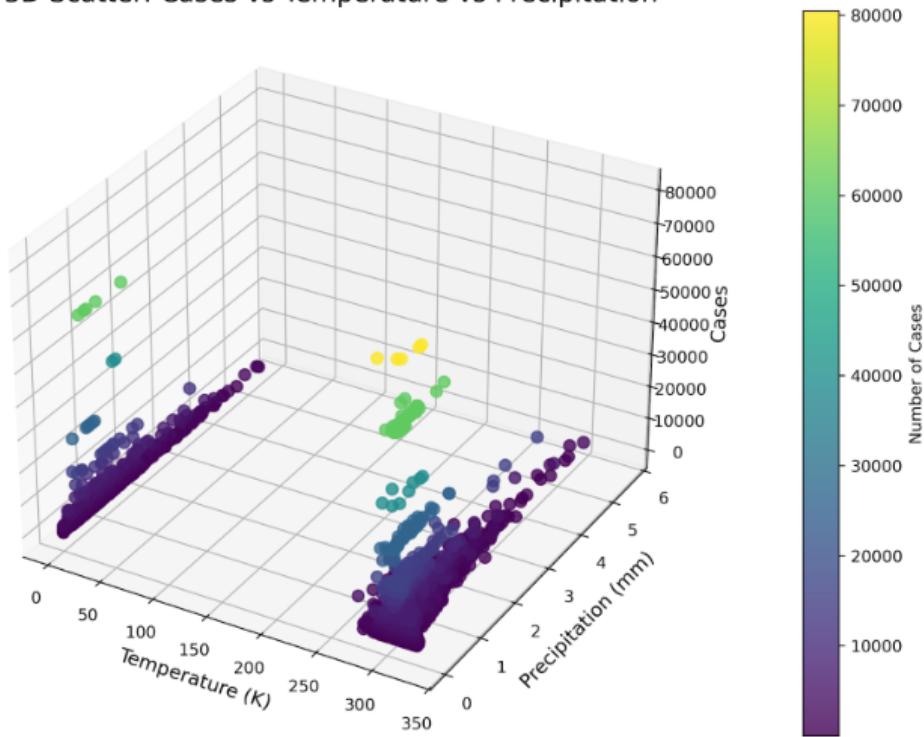


**Figure 15:** Spatial distribution of Malaria cases in India in 2022.

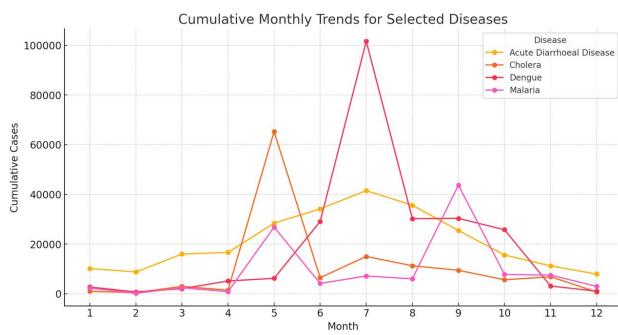


**Figure 16:** Spatial distribution of Malaria cases in India across selected years.

3D Scatter: Cases vs Temperature vs Precipitation

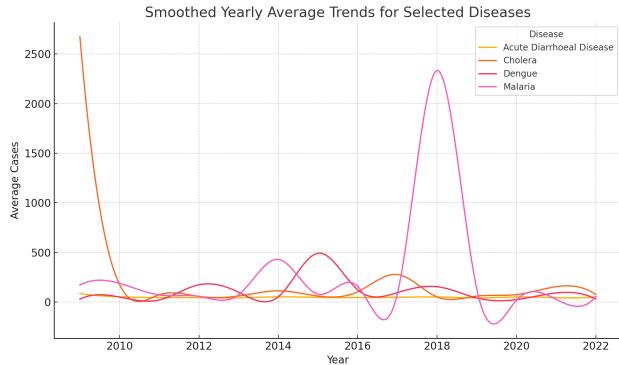


**Figure 17:** 3D Scatter Plot of Disease Cases vs. Temperature and Precipitation. The color scale represents the number of cases, with warmer colors indicating higher case counts.

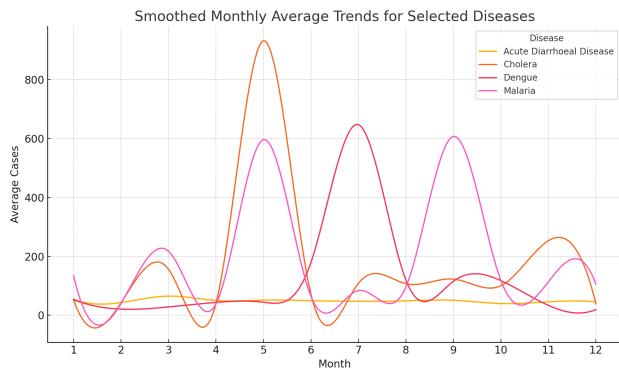


**Figure 18:** cumulative monthly trend for selective diseases

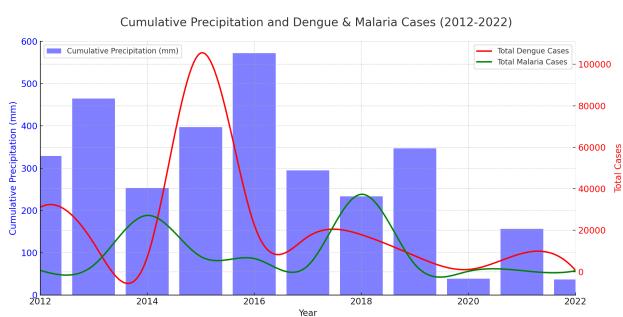
## Climate Health Nexus



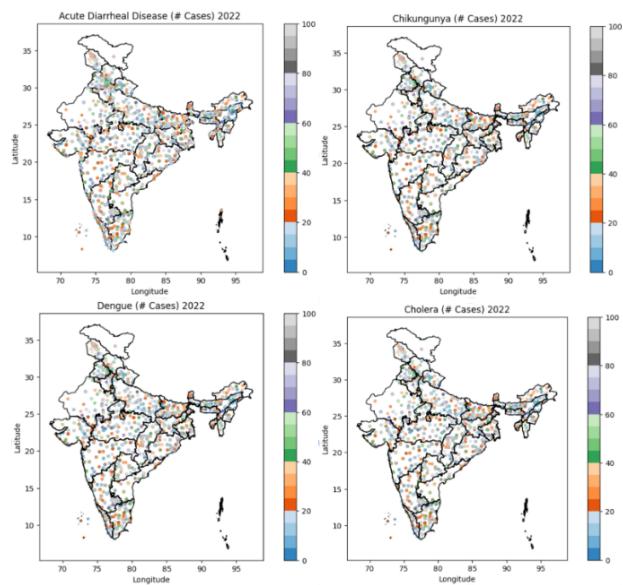
**Figure 19:** Smoothed yearly average trend for selective diseases



**Figure 20:** Smoothed months average trend for selective diseases



**Figure 21:** Cumulative Precipitation for Dengue and Malaria Cases(2012-2022)



**Figure 22:** all four diseases.