Real-Time Mapping of Epidemic Spread

A PROJECT REPORT

Submitted by,

SI.No	Roll Number	Student Name
1	20211CSE0141	Ayan Bhattacharya
2	20211CSE0094	Raj Kumar Chandra
3	20211CSE0186	Kishnaram Jaswanth
4	20211CSE0050	Swastik.R.Gurung
5	20211CSE0047	Shovan Patra

Under the guidance of,

Prof. Jayachandran Arumugam

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING,

At

PRESIDENCY UNIVERSITY

BENGALURU

DECEMBER 2025

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report **"Real-Time Mapping of Epidemic Spread"** being submitted by "Ayan Bhattacharya, Raj Kumar Chanda, Kishnaram Jaswanth, Swastik Gurung, Shovan Patra" bearing roll number(s) "20211CSE0141, 20211CSE0094, 20211CSE0186, 20211CSE0050, 20211CSE0047" in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

Prof.Jayachandran Arumugam
DESIGNATION
School of CSE&IS
Presidency University

Dr. Blessed Prince P/Dr. Robin Rohit/Dr. Asif Mohammed H.B
DESIGNATION & HoD
School of CSE&IS
Presidency University

Dr. L. SHAKKEERA
Associate Dean
School of CSE
Presidency University

Dr. MYDHILI NAIR
Associate Dean
School of CSE
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-VC School of Engineering
Dean -School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **Real-Time Mapping of Epidemic Spread** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **Prof.Jayachandran Arumugam , DESIGNATION, School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

SI.No	Roll Number	Student Name	Signature
1	20211CSE0141	Ayan Bhattacharya	
2	20211CSE0094	Raj Kumar Chandra	
3	20211CSE0186	Kishnaram Jaswanth	
4	20211CSE0050	Swastik.R.Gurung	
5	20211CSE0047	Shovan Patra	

ABSTRACT

This project presents a novel epidemic tracking application that overcomes the limitations of traditional systems like HealthMap and MedISys by integrating Generative AI (GenAI) with modern web technologies for advanced information extraction and dynamic visualization. Unlike existing systems, which rely on predefined sources and structured data, our solution processes real-time unstructured data from platforms like Google News, WHO, and academic journals using advanced NLP techniques such as Named Entity Recognition (NER), text summarization, and topic modeling. The extracted data is then visualized on an interactive map, built with React and Leaflet, offering real-time outbreak patterns and trends with dynamic markers and overlays.

This system provides a scalable and adaptable solution for epidemic monitoring, overcoming static, structured data limitations. It can dynamically ingest and synthesize data from multiple sources, including multilingual ones, ensuring global applicability. Future improvements will focus on predictive analytics and expanding the data collection scope. By combining AI and visualization, our tool enhances real-time, global outbreak monitoring, offering a significant leap forward compared to traditional epidemic tracking systems.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair,** School of Computer Science Engineering & Information Science, Presidency University, and **Dr. Asif Mohammed H.B**, Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Prof.Jayachandran Arumugam**, School of Computer Science Engineering & Information Science, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar and Mr. Md Zia Ur Rahman,** Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Ayan Bhattacharya,
Raj Kumar Chanda,
Kishnaram Jaswanth,
Shovan Patra,
Swastik Ratna Gurung

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

1.1 **Overview of Epidemics and Real-Time Tracking**

1.1.1 **The Need for Advanced Surveillance Systems**

Epidemics have historically disrupted societies, causing widespread illness, loss of life, and economic instability. The swift nature of disease outbreaks often overwhelms traditional surveillance and response systems, leaving communities vulnerable. In this context, the importance of real-time epidemic tracking has grown significantly. Advanced systems that collect, process, and display real-time data can bridge the critical gap between detection and intervention, enabling stakeholders to act promptly and effectively.

1.1.2 **The Role of Technology in Modern Epidemic Management**

Technological innovations have redefined epidemic surveillance by enabling real-time data collection and visualization. Tools like artificial intelligence, machine learning, and geospatial mapping have made it possible to monitor disease transmission patterns with unprecedented precision. These advancements help identify hotspots, predict outbreak trends, and inform targeted response strategies, ultimately saving lives and resources.

1.2 **Technological Advancements in Epidemic Mapping**

The integration of data from diverse sources—such as public health APIs, mobility patterns, and demographic statistics—has enhanced the capabilities of epidemic management platforms. Real-time mapping tools leverage these datasets to provide accurate, location-specific insights. By using advanced algorithms and visualization techniques, these systems not only track disease progression but also forecast potential outbreaks, enabling proactive measures. This project employs cutting-edge technologies such as artificial intelligence for predictive modeling, Leaflet.js for interactive maps, and MongoDB for robust data handling. Together, these tools form a comprehensive solution to address the complex challenges posed by epidemic monitoring and control.

1.3 **Project Significance and Objectives**

The "Real-Time Mapping of Epidemic Spread" project is a crucial step toward enhancing global epidemic preparedness. By developing a platform that integrates real-time data collection, visualization, and predictive analytics, the project seeks to:

- Provide stakeholders with actionable insights through an interactive mapping interface.
- Enable timely interventions by issuing real-time alerts tailored to specific regions and user locations.
- Support public health officials in resource allocation and policy-making.
- Improve public awareness and collaboration through accessible and transparent data sharing.

CHAPTER-2

LITERATURE SURVEY

Real-Time Epidemic Forecasting for Pandemic Preparedness- Data latency and inaccuracy when mobility and genomic data are not updated promptly.

• Quantifying SARS-CoV-2 Transmission and Epidemic Control with Digital Contact Tracing- Privacy concerns lead to low adoption rates and insufficient data for modeling.

• Modeling Epidemics with Compartmental Models and Bayesian Networks - Compartmental models may oversimplify complex population interactions leading to inaccurate forecasts.

• The Use of Mobile Phone Data for the Real-Time Tracking of Infectious DiseasesMobile phone data can be biased, missing vulnerable populations without access.

• Graph Attention-Based Spatial-Temporal Networks for Epidemic Forecasting- Requires large, high-quality datasets; sparse data can hinder performance.

• Real-Time Bayesian Estimation of Epidemic Parameters Using Case Data - Computationally expensive when dealing with large-scale real-time data.

• The Role of Big Data and Machine Learning in Predicting and Managing Epidemics - Models may overfit with noisy or heterogeneous data, reducing generalizability.

• OutbreakFlow: Model-Based Bayesian Inference of Disease Outbreak Dynamics - Bias may occur if prior distributions are inaccurately set or too restrictive.

• Deep Learning for Epidemic Spread Prediction Using Spatio-Temporal Data- Requires large volumes of labeled data which may not be available for all regions or diseases.

• Using Satellite Imagery for Predicting Epidemic Outbreaks - Limited resolution or frequency makes it difficult to capture real-time changes or smaller-scale outbreaks.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

1. Quantifying SARS-CoV-2 Transmission and Epidemic Control with Digital Contact Tracing Authors: Luca Ferretti, Chris Wymant, Michelle Kendall, link: https://pubmed.ncbi.nlm.nih.gov/32234805/
2. Bayesian compartmental model for an infectious disease with dynamic states of infection Authors: Marie V Ozanne , Grant D Brown , Jacob J Oleson Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC6752225/
3. The Use of Mobile Phone Data for the Real-Time Tracking of Infectious Diseases Authors: Ymir Vigfusson, Thorgeir A Karlsson, Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC8017972/
4. Modeling epidemic dynamics using Graph Attention based Spatial Temporal networks Authors: Xiaofeng Zhu,Yi Zhang, Link: https://journals.plos.org/plosone/article?id=10.1371/
5. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases Authors: Luís M A Bettencourt, Ruy M Ribeiro, Link: https://pmc.ncbi.nlm.nih.gov/articles/
6. Application of big data and artificial intelligence in epidemic surveillance and containment Authors: Zengtao Jiao, Hanran Ji , Link: https://www.sciencedirect.com/science/article/pii/
7. OutbreakFlow: Model-based Bayesian inference of disease outbreak dynamics with invertible neural networks and its application to the COVID-19 pandemics in Germany Authors: Stefan T. Radev ,Frederik Graw,Simiao Chen, Link: https://journals.plos.org/ploscompbiol/
8. Forecasting infections with spatio-temporal graph neural networks: a case study of the Dutch SARS-CoV-2 spread Authors: V. Maxime Croft, Link: https://www.frontiersin.org/journals/physics/articles/
9. Using Satellite Images of Environmental Changes to Predict Infectious Disease Outbreaks Authors: Timothy E Ford , Rita R Colwell , Link: https://pmc.ncbi.nlm.nih.gov/articles/
10. REAL-TIME MECHANISTIC BAYESIAN FORECASTS OF COVID-19 MORTALITY Authors: GRAHAM C GIBSON, Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC7781348/#ABS1

CHAPTER-4

PROPOSED MOTHODOLOGY

- **Real-Time Epidemic Tracking Using Flask and Leaflet.js:** Utilized Flask for a scalable backend and Leaflet.js for interactive map visualization of disease spread.
- **Automated News Article Classification for Disease Tracking:** Implemented RSS feed parsing for data collection and GPT-3.5 to classify articles as relevant or not.
- **Named Entity Recognition for Epidemic Data Extraction:** Used Kor for extracting key entities like date, disease, location, severity, species, and summary.
- **Visualizing Epidemic Data Using React and Axios:** Developed a frontend using React that retrieves data from the Flask backend through Axios and displays it on an interactive map.
- **Enhancing User Interaction with Map-Based Disease Insights:** Created tooltips that display crucial information when hovering over markers, providing users with real-time updates.
- **Potential Improvements in Data Classification:** Highlighted the need for more accurate classification techniques to avoid false positives in identifying relevant articles.
- **Dynamic Data Handling with Flask:** Leveraged Flask to handle dynamic data input and output, ensuring real-time updates for users.
- **Mapping Disease Severity with Color-Coded Markers:** Implemented a color-coded system to indicate the severity of outbreaks, improving data visualization and user experience.

CHAPTER-5

OBJECTIVES

- Collect real-time epidemic data from various sources, including public health APIs and local reports, focusing on India.
- Display the collected data on an interactive map to visualize affected regions and severity levels effectively.
- Provide real-time risk alerts tailored to users based on their location and the latest epidemic data.
- Integrate multiple data sources to ensure accurate and timely updates through automated data fetching.
- Offer an intuitive and accessible user interface to facilitate easy navigation and data filtering for users.
- Implement real-time data integration from mobility sources to enhance epidemic forecasting accuracy.
- Improve public health communication by providing visualizations and dashboards for real-time data sharing and accessibility.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

Architecture Overview:

- The system follows a client-server architecture, with a Flask backend for data processing and a React-based frontend for user interaction.
- A RESTful API built using Flask handles communication between the backend and the frontend, enabling data exchange in JSON format.

Backend Design:

- Data Collection:
  - Data is collected from various sources such as RSS feeds of news and journal websites.
  - An RSS feed parser extracts articles containing predefined search terms related to disease outbreaks.
- Data Filtering:
  - OpenAI's GPT-3.5 API is employed to determine the relevance of articles. Each article is classified as "Yes" or "No" based on whether it discusses a recent disease spread.
- Entity Extraction:
  - The Kor library is used for Named Entity Recognition (NER) to extract key entities such as disease name, location, date, severity, species, and a summary.
- Data Transmission:
  - Extracted data is converted into JSON format and transmitted to the frontend via a Flask API endpoint.

Frontend Implementation:

- Built using React, the frontend fetches data from the backend using Axios.
- The Leaflet.js library is integrated for mapping and visualization:
  - Disease locations are marked with colored circles based on severity (e.g., red for high severity, orange for medium, and green for low).
  - Tooltips provide additional information, including the date and a summary of the disease outbreak.

Technologies Used:

- Backend: Flask, RSS feed parser, OpenAI API, Kor library.
- Frontend: React, Axios, Leaflet.js.
- Database: Data is temporarily stored and processed in-memory during execution

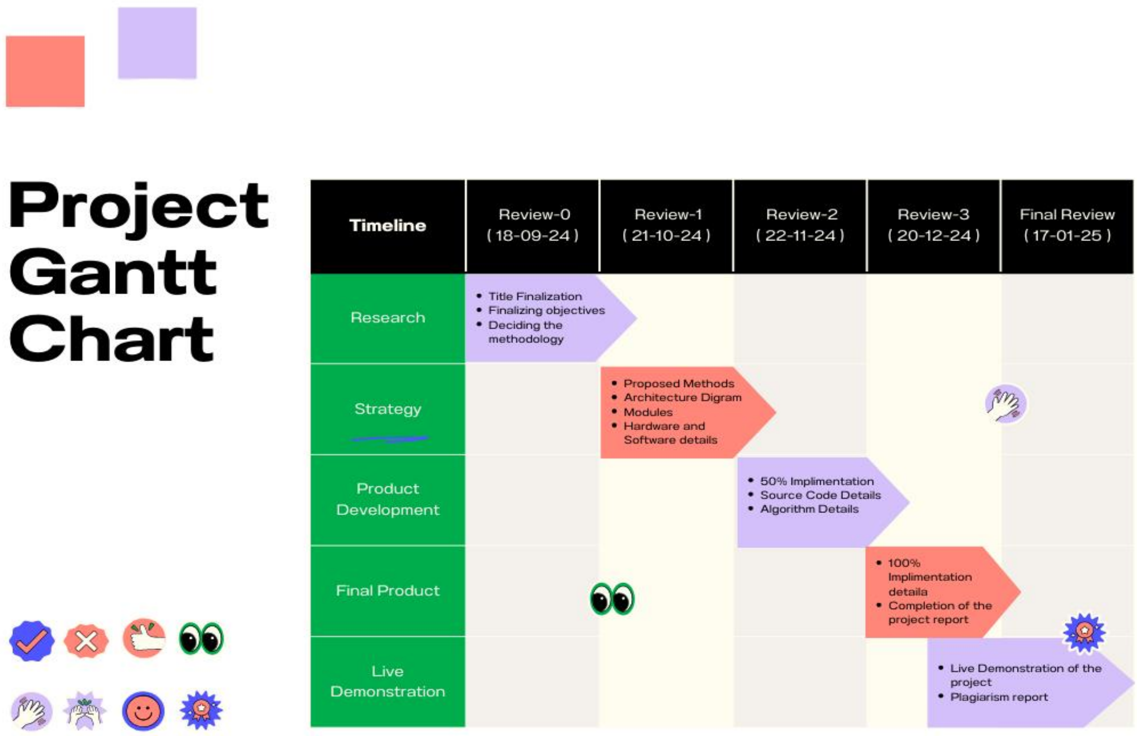
Integration:

- Smooth integration between the backend and frontend was achieved through consistent data structures (JSON format)

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)



CHAPTER-8

OUTCOMES

- Automated data collection from various online sources through RSS feed parsing.

- Utilization of GPT-3.5 for classifying articles related to disease outbreaks.

- Extraction of key details such as date, disease, location, severity, species, and summary using Kor for NER.

- Conversion of processed data into JSON format and serving it through a Flask backend.

- Visualization of disease spread on an interactive map using Leaflet.

- Color-coded markers to indicate severity levels and tooltips displaying essential details on hover.

- Real-time insights into ongoing disease outbreaks and their geographical impact.

- Enhanced public awareness and decision-making support for health organizations.

- Identified limitation of false positives due to general-purpose AI models.

CHAPTER-9

RESULTS AND DISCUSSIONS

The Real-Time Mapping of Epidemic Spread system collects data from news and journal websites using RSS feed parsing. Relevant articles are identified using GPT-3.5, and Named Entity Recognition (NER) is performed using Kor to extract key details like date, disease, location, severity, species, and summary. The processed data is converted to JSON and served through a Flask backend. The frontend, built with React, retrieves the data via Axios and uses Leaflet to display interactive maps with color-coded markers based on disease severity. Tooltips with date and summary appear when users hover over the markers.

The project offers real-time insights into disease outbreaks and helps visualize their geographical spread, supporting public awareness and decision-making by health organizations. However, some limitations affect its accuracy. The GPT-3.5 model occasionally misclassifies articles. For instance, an article titled "How COVID Affected Mumbai in 2022" might be tagged as relevant despite not discussing current outbreaks. This results in false positives, reducing the system's reliability.

To address this, a specialized NLP model trained to identify ongoing outbreaks could be integrated. Such a model would enhance the system’s precision by better distinguishing between past events and current reports. Additionally, the system relies solely on public data, which can be inaccurate or delayed. Introducing a secure portal for verified health workers to submit reports could improve the reliability and timeliness of the data. Identity verification and a review process for submissions would ensure data quality, making the system more robust and actionable.

CHAPTER-10

CONCLUSION

The project demonstrates the potential of leveraging modern web technologies and AI to improve disease surveillance. By automating the process of collecting, processing, and visualizing data from various sources, the project provides timely information that can help in mitigating the impact of disease outbreaks. The interactive mapping feature makes it easy for users to identify affected areas and assess the severity of outbreaks.

However, the project also highlights the challenges of using general-purpose AI models for domain-specific tasks. The occurrence of false positives indicates the need for more specialized tools to improve accuracy. Future enhancements, such as the inclusion of a health worker portal for direct reporting and the integration of task-specific NLP models, could significantly improve the system's reliability and utility.

Overall, this project lays a strong foundation for further development in real-time epidemic tracking systems. It has the potential to be a valuable tool for public health authorities, researchers, and the general public, contributing to better preparedness and response in managing disease outbreaks.

REFERENCES

1. Quantifying SARS-CoV-2 Transmission and Epidemic Control with Digital Contact Tracing Authors: Luca Ferretti, Chris Wymant, Michelle Kendall, link: https://pubmed.ncbi.nlm.nih.gov/32234805/
2. Bayesian compartmental model for an infectious disease with dynamic states of infection Authors: Marie V Ozanne , Grant D Brown , Jacob J Oleson Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC6752225/
3. The Use of Mobile Phone Data for the Real-Time Tracking of Infectious Diseases Authors: Ymir Vigfusson, Thorgeir A Karlsson, Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC8017972/
4. Modeling epidemic dynamics using Graph Attention based Spatial Temporal networks Authors: Xiaofeng Zhu,Yi Zhang, Link: https://journals.plos.org/plosone/article?id=10.1371/
5. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases Authors: Luís M A Bettencourt, Ruy M Ribeiro, Link: https://pmc.ncbi.nlm.nih.gov/articles/
6. Application of big data and artificial intelligence in epidemic surveillance and containment Authors: Zengtao Jiao, Hanran Ji , Link: https://www.sciencedirect.com/science/article/pii/
7. OutbreakFlow: Model-based Bayesian inference of disease outbreak dynamics with invertible neural networks and its application to the COVID-19 pandemics in Germany Authors: Stefan T. Radev ,Frederik Graw,Simiao Chen, Link: https://journals.plos.org/ploscompbiol/
8. Forecasting infections with spatio-temporal graph neural networks: a case study of the Dutch SARS-CoV-2 spread Authors: V. Maxime Croft, Link: https://www.frontiersin.org/journals/physics/articles/
9. Using Satellite Images of Environmental Changes to Predict Infectious Disease Outbreaks Authors: Timothy E Ford , Rita R Colwell , Link: https://pmc.ncbi.nlm.nih.gov/articles/
10. REAL-TIME MECHANISTIC BAYESIAN FORECASTS OF COVID-19 MORTALITY Authors: GRAHAM C GIBSON, Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC7781348/#ABS1
Previous Projects
Real-Time Mapping of Epidemic Spreads Link: https://github.com/ShreyaGupta08/Epidemic-Spread-SIH

APPENDIX-A

PSUEDOCODE

1.Data Collection and Processing

```
START Backend System

SET sources = ["News Websites", "Journal Websites"]

FUNCTION fetch_and_process_articles(sources)
    FOR each source IN sources
        articles = RSSFeedParser(source)
        FOR each article IN articles
            IF GPT3.5("Does this article talk about recent disease
spread?", article) == "Yes"
                entities = ExtractEntitiesWithKor(article)
                structured_data = {
                    "date": entities.date,
                    "disease": entities.disease,
                    "location": entities.location,
                    "severity": entities.severity,
                    "species": entities.species,
                    "summary": entities.summary
                }
                StoreDataAsJSON(structured_data)
            END IF
        END FOR
    END FOR
END FUNCTION

FUNCTION ExtractEntitiesWithKor(article)
    RETURN Kor(article).extract_entities()
END FUNCTION

FUNCTION StoreDataAsJSON(data)
    CONVERT data TO JSON
    SEND JSON TO "http://<flask_server_ip>:<port>"
END FUNCTION
```

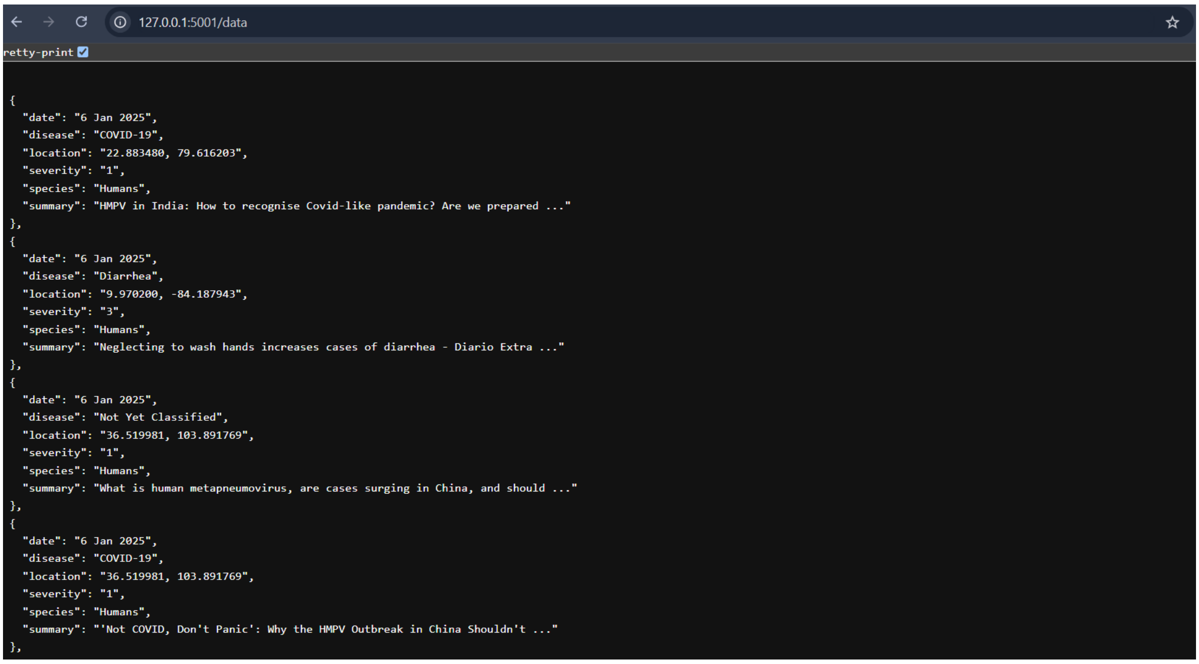
2.Frontend: Data Retrieval and Visualization

```
START Frontend System

FUNCTION collect_data_from_backend()
    data = HTTPGetRequest("http://<flask_server_ip>:<port>")
    RETURN data
END FUNCTION

FUNCTION display_data_on_map(data)
    FOR each entry IN data
        SET severityColor = DetermineColor(entry.severity)
        DrawCircleOnMap(entry.location, severityColor)
        AttachTooltip(entry.location, entry.date, entry.summary)
    END FOR
```

```
END FUNCTION

FUNCTION DetermineColor(severity)
    IF severity == "High"
        RETURN "Red"
    ELSE IF severity == "Medium"
        RETURN "Orange"
    ELSE
        RETURN "Green"
    END IF
END FUNCTION

FUNCTION AttachTooltip(location, date, summary)
    tooltip = "Date: " + date + "\nSummary: " + summary
    AttachTooltipToMap(location, tooltip)
END FUNCTION

CALL collect_data_from_backend()
CALL display_data_on_map(data)
```

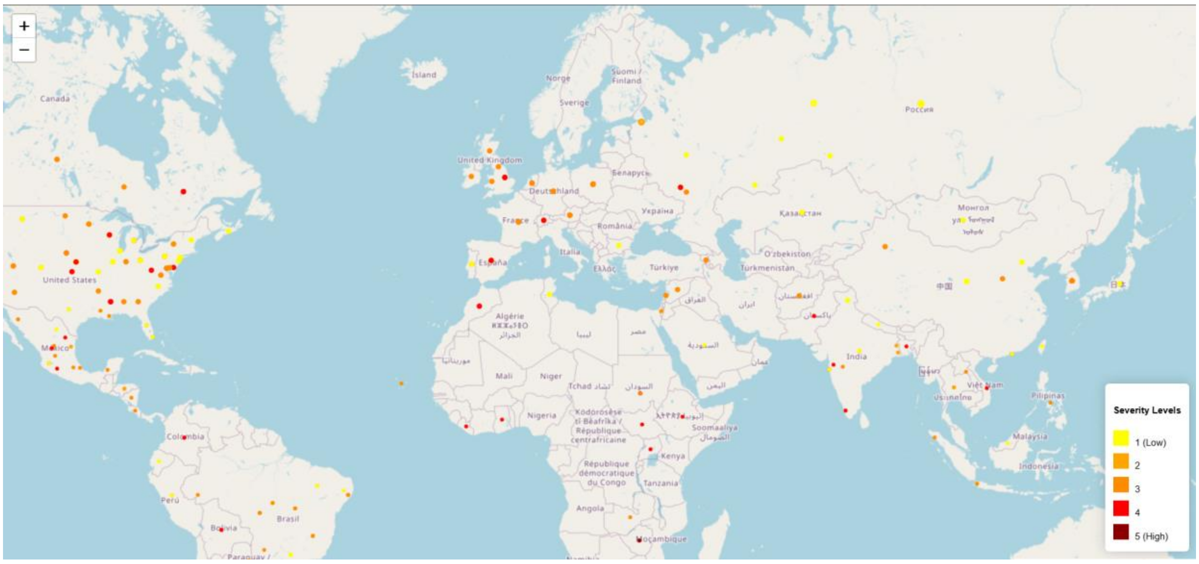
APPENDIX-B

SCREENSHOTS

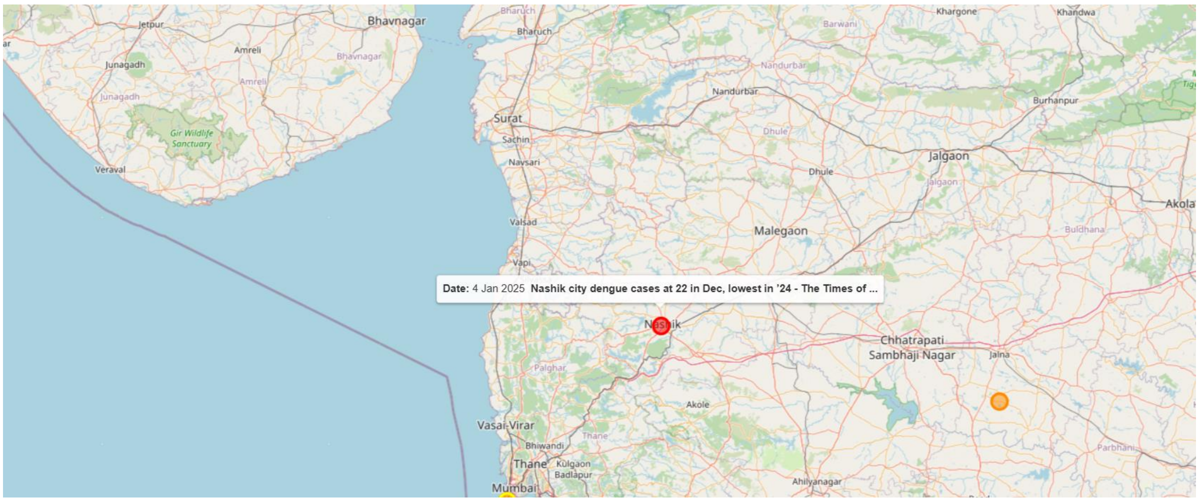


{
  "date": "6 Jan 2025",
  "disease": "COVID-19",
  "location": "22.883480, 79.616203",
  "severity": "1",
  "species": "Humans",
  "summary": "HMPV in India: How to recognise Covid-like pandemic? Are we prepared ..."
},
{
  "date": "6 Jan 2025",
  "disease": "Diarrhea",
  "location": "9.970200, -84.187943",
  "severity": "3",
  "species": "Humans",
  "summary": "Neglecting to wash hands increases cases of diarrhea - Diario Extra ..."
},
{
  "date": "6 Jan 2025",
  "disease": "Not Yet Classified",
  "location": "36.519981, 103.891769",
  "severity": "1",
  "species": "Humans",
  "summary": "What is human metapneumovirus, are cases surging in China, and should ..."
},
{
  "date": "6 Jan 2025",
  "disease": "COVID-19",
  "location": "36.519981, 103.891769",
  "severity": "1",
  "species": "Humans",
  "summary": "'Not COVID, Don't Panic': Why the HMPV Outbreak in China Shouldn't ..."
},

Data Extraction



Final Output



Tool-Tip Function on hover for more information

APPENDIX-C
ENCLOSURES

1. Journal publication/Conference Paper Presented Certificates of all students.

2. Include certificate(s) of any Achievement/Award won in any project-related event.

3. Similarity Index / Plagiarism Check report clearly showing the Percentage (%). No need for a page-wise explanation.

4. Details of mapping the project with the Sustainable Development Goals (SDGs).