PRESIDENCY UNIVERSITY

Private University Estd. in Karnataka State by Act No. 41 of 2013

BANGALORE

A Project Report

On

Real Time Mapping of Epidemic Spread

Batch Details

Sl. No.	Roll Number	Student Name
1	20211CSD0038	B.N BHAVANA
2	20211CSD0005	PRARTHANA S.P
3	20211CSD0042	ULLAS GOWDA M
4	20211CSD0150	SANCHIT A

School of Computer Science,

Presidency University, Bengaluru.

Under the guidance of,

Prof. Tintu Vijayan

School of Computer Science,

Presidency University, Bengaluru

CONTENTS

1. *INTRODUCTION*

The rapid spread of infectious diseases like dengue, Zika, and influenza presents significant public health challenges globally. Outbreaks of such diseases often escalate quickly due to the fast transmission rates, especially in densely populated regions and areas with inadequate healthcare infrastructure. Timely intervention and efficient resource allocation are critical in mitigating the impact of such epidemics. Traditional methods of epidemic management, which often rely on delayed data analysis, can result in reactive rather than proactive responses. This is where real-time mapping of epidemic spread becomes invaluable.

Real-time mapping integrates multiple layers of data—including epidemiological reports, environmental factors, and mobility patterns—to provide a dynamic, continuously updated picture of how diseases are spreading across regions. The goal is to give public health officials and policymakers the ability to visualize disease transmission patterns in real time, identify hotspots, and implement targeted control measures. With the rise of Geographic Information Systems (GIS) and advancements in machine learning, the tools for epidemic forecasting have evolved significantly. Real-time mapping also allows for the incorporation of multiple data sources, from traditional case reporting to social media trends and environmental sensors, which can enhance predictive accuracy.

This project aims to develop a robust, scalable real-time mapping system that can be used for various epidemics, improving public health preparedness. Through the combination of spatial and temporal data, it seeks to highlight high-risk areas and guide timely interventions, ensuring that healthcare resources are directed where they are most needed. The mapping system will be designed with a focus on user-friendly visualization, allowing health authorities to monitor and respond to emerging outbreaks efficiently. By integrating real-time data into predictive models, the project will enable a proactive, data-driven approach to epidemic control.

Real-time mapping of epidemic spread is an essential aspect of contemporary public health, utilizing advanced technologies such as Geographic Information Systems (GIS), data analytics, and mobile applications to effectively monitor and respond to infectious disease outbreaks. This approach allows health authorities to visualize the geographic distribution of cases, facilitating early detection of outbreaks and enabling timely interventions that can prevent widespread transmission. By providing insights into population density, healthcare infrastructure, and environmental factors, real-time mapping aids in optimizing resource allocation, ensuring that medical personnel, vaccines, and treatment facilities are directed to areas most in need.

Moreover, interactive maps and dashboards enhance public awareness by delivering timely information about ongoing outbreaks, empowering individuals to take appropriate precautions and reducing panic. The integration of data from diverse sources—such as hospitals, laboratories, and public health agencies—enables a comprehensive understanding of disease dynamics. However, challenges persist, including ensuring data quality and availability, addressing privacy concerns related to sensitive health information, and adapting to the dynamic nature of epidemics. As recent global health crises have underscored, the ability to map and analyze epidemic spread in real time is vital for mitigating the impact of emerging infectious diseases, ultimately improving public health outcomes and resilience against future outbreaks.

2. LITERATURE REVIEW

- Epidemiological Models and Surveillance Data

Epidemiological models have long been the cornerstone of disease surveillance and control. Traditional models, such as the Susceptible-Infectious-Recovered (SIR) framework, rely heavily on historical data and time series analysis. These models focus on understanding how diseases progress over time, often using confirmed case reports from health agencies to estimate infection rates. While these models are effective in tracking ongoing outbreaks, they tend to lag in early-stage outbreaks, where real-time data is scarce. The unpredictable nature of disease transmission, especially in low-resource settings, introduces significant uncertainties. Early stages of outbreaks like H5N1 avian influenza in Vietnam, for example, demonstrated the limitations of models based purely on temporal data, as they failed to capture the complexity of the disease's spatial distribution.

- Incorporating Spatial Dynamics

One of the critical advancements in recent epidemic forecasting is the integration of spatial data into traditional epidemiological models. Geographic Information Systems (GIS) have emerged as powerful tools for visualizing and analyzing the geographical spread of diseases. GIS allows researchers to map the distribution of confirmed cases and overlay it with various environmental and social factors, such as population density, rainfall patterns, and proximity to bodies of water. By incorporating these variables, models become more comprehensive, allowing for more accurate predictions of where outbreaks are likely to intensify.
Moreover, participatory mapping approaches that incorporate community knowledge into GIS have been used to assess dengue risk in specific areas. In Malaysia, for example, researchers employed local community knowledge to map areas where mosquito breeding was most prevalent, significantly improving the accuracy of spatial risk models. Similarly, advanced techniques like Geographically Weighted Regression (GWR) and the Local Index of Spatial Autocorrelation (LISA) have enabled researchers to evaluate how socioeconomic factors and environmental variables correlate with disease incidence.

- Real-Time Forecasting

The ability to forecast outbreaks in real time is critical to effective epidemic management. Bayesian estimation methods, for example, are commonly used to calculate the effective reproductive number (R), which indicates the number of new cases generated by each infected person. This dynamic estimate helps public health officials understand how an outbreak is evolving and when it is likely to peak. Such real-time estimates are crucial for allocating resources effectively and ensuring timely interventions.
Studies have also explored the use of unconventional data sources for early outbreak detection. For example, real-time analysis of Google search trends and phone call logs has been shown to predict influenza outbreaks several weeks before official health alerts are issued. This approach relies on analyzing spikes in search queries for terms like "flu symptoms," providing an early indication of disease spread. Similarly, information theory has been used to optimize epidemic forecasting models, focusing on adaptive window sizing to capture real-time changes in disease dynamics.

Existing methods and it’s advantages and disadvantages

1. Geographic Information Systems (GIS)
Advantages:

Spatial Analysis: GIS allows for detailed spatial analysis, enabling the visualization of epidemic spread across different geographic areas.
Layered Data: It can integrate multiple data layers (e.g., population density, healthcare facilities) for comprehensive insights.

Limitations:

Data Quality: The effectiveness of GIS depends on the accuracy and timeliness of the input data, which can vary significantly.
Complexity: Requires specialized skills to analyze and interpret GIS data, potentially limiting accessibility for non-experts.

2. Mobile Health Applications
Advantages:

Real-Time Reporting: Users can report symptoms and outbreaks in real time, providing immediate data for mapping.
Wide Reach: Mobile apps can reach a broad audience, facilitating widespread participation in data collection.

Limitations:

Digital Divide: Not everyone has access to smartphones or the internet, leading to potential data gaps.
User Bias: Self-reported data may be biased or inaccurate, affecting overall reliability.

3. Epidemiological Modeling
Advantages:

Predictive Analytics: Models can predict future outbreak scenarios based on current data, aiding in proactive planning.
Scenario Testing: Allows for the simulation of various intervention strategies to assess their potential impact.

Limitations:

Assumptions and Simplifications: Models often rely on assumptions that may not hold true in real-world scenarios, leading to inaccurate predictions.
Data Dependency: The accuracy of models is heavily reliant on the quality of data input, which can be inconsistent.

4. Social Media Monitoring
Advantages:

Rapid Data Collection: Social media platforms can provide real-time insights into public sentiment and potential outbreak reports.
Broad Coverage: Can capture data from diverse populations, offering a more comprehensive view of public health trends.

Limitations:

Noise and Misinformation: High volumes of irrelevant data and misinformation can obscure valuable insights.
Privacy Concerns: Analyzing public posts raises ethical issues regarding privacy and consent.

5. Remote Sensing Technologies
Advantages:

Wide-Area Surveillance: Satellite imagery can monitor large geographic areas, providing valuable data on environmental factors affecting disease spread.
Objective Data Collection: Reduces human bias in data collection, offering objective insights.

Limitations:

High Cost: Accessing and processing satellite data can be expensive and resource-intensive.
Limited Resolution: The resolution of satellite images may not always be sufficient for detailed epidemiological analysis.

6. Wearable Health Devices
Advantages:

Continuous Monitoring: Wearable devices can track health metrics continuously, providing real-time data on population health.
Personalized Data: Offers insights into individual health trends, which can contribute to broader epidemiological understanding.

Limitations:

Data Privacy: Concerns over the security and privacy of health data collected by wearables can hinder widespread adoption.
Limited Adoption: Not everyone uses wearable technology, leading to potential biases in the data collected.

7. Public Health Dashboards
Advantages:

Centralized Information: Dashboards aggregate data from various sources, providing a comprehensive overview of epidemic status.
User-Friendly: Often designed for ease of use, making them accessible to the general public and decision-makers.

Limitations:

Data Lag: There can be delays in data updating, leading to outdated information being presented.
Over-Simplification: Dashboards may oversimplify complex data, potentially obscuring critical insights.

8. Crowdsourcing Platforms
Advantages:

Community Engagement: Encourages public participation in data collection, fostering community involvement in health monitoring.
Diverse Data Sources: Can gather information from various demographics, enhancing data richness.

Limitations:

Data Reliability: Crowdsourced data may lack verification, leading to questions about its accuracy.
Inconsistent Participation: Participation levels can fluctuate, resulting in uneven data coverage.

9. Contact Tracing Applications
Advantages:

Immediate Alerts: Can notify users if they have been in contact with someone who tested positive for an infectious disease.
Data-Driven Insights: Provides valuable data for understanding transmission patterns.

Limitations:

Privacy Concerns: Users may be reluctant to share location data, fearing surveillance or misuse of information.
Limited Effectiveness: Relies on widespread adoption to be effective; low uptake can diminish its impact.

10. Traditional Surveillance Systems
Advantages:

Established Protocols: Well-defined methods for data collection and reporting, often backed by public health authorities.
Comprehensive Data: Can provide extensive historical data for analysis.

Limitations:

Slow Response: Traditional systems may not provide real-time data, delaying response efforts.
Resource Intensive: Often requires significant manpower and resources to maintain and operate effectively.

Challenges and Opportunities

Despite the advancements in real-time epidemic mapping, several challenges persist. One of the main challenges is the availability and quality of real-time data. In many regions, particularly in low-resource settings, data may be incomplete or delayed, making it difficult to generate accurate predictions. Additionally, the complexity of integrating diverse data sources—ranging from epidemiological reports to social media trends—requires sophisticated data-cleaning techniques and robust algorithms. Model complexity also poses a challenge, as more advanced models may be difficult to interpret and may require significant computational resources.
Future research should focus on addressing these challenges by developing more adaptable models that can handle diverse data inputs and account for spatial heterogeneity. There is also a need for interdisciplinary collaboration between epidemiologists, data scientists, and public health officials to ensure that models are both scientifically rigorous and practically useful.

3. OBJECTIVES

The primary objectives of this project are as follows:

- Develop a Real-Time Mapping System: The project aims to build a robust, real-time mapping system that visualizes the geographical spread of an epidemic as it unfolds. This system will leverage GIS technology to create dynamic maps that can be updated with real-time epidemiological data.

- Identify High-Risk Areas: By analyzing both spatial and temporal data, the project seeks to identify high-risk areas where outbreaks are likely to intensify. These areas will be flagged for public health officials, enabling targeted interventions to prevent further spread.
- Provide Timely Insights: The project will focus on providing timely insights to guide public health interventions. This includes predicting the future trajectory of the epidemic, highlighting regions that require immediate attention, and informing resource allocation decisions.
- Integrate Real-Time Data: The mapping system will incorporate various real-time data sources, including confirmed cases, environmental conditions, and mobility patterns. This integration will enhance the predictive power of the model and allow for dynamic updates as new data becomes available.

By achieving these objectives, the project will contribute to more effective public health management and improve the ability of authorities to respond quickly to emerging outbreaks. The system will be designed for scalability, allowing it to be used for various diseases and in different geographical contexts.

<u>EXPERIMENTAL DETAILS/METHDOLOGY</u>

Data Analytics Tools:

Examples: Python (with libraries like Pandas, NumPy), R, MATLAB
Function: For data processing, statistical analysis, and modeling of epidemic trends.

Machine Learning Frameworks:

Examples: TensorFlow, Scikit-learn, Keras
Function: For developing predictive models to forecast epidemic spread based on historical and real-time data.

Visualization Tools:

Examples: Tableau
Function: For creating interactive dashboards and visual representations of epidemic data.

<u>4. METHODOLOGY</u>

<u>- DESIGN PROCEDURE</u>

- <u>Data Collection</u>

  The first step in developing the real-time mapping system is to gather relevant epidemiological data. This includes confirmed case reports, suspected cases, and mortality data from public health agencies, hospitals, and surveillance networks. Data on environmental factors, such as temperature, humidity, and rainfall, will also be collected to account for variables that may influence disease transmission. In addition, mobility data from sources like public transportation systems and mobile phones will be incorporated to track how human movement patterns contribute to the spread of the disease.

- <u>Data Preprocessing</u>

  Once the data is collected, it will undergo rigorous preprocessing to ensure it is suitable for analysis. This involves cleaning the data by addressing missing values, removing inconsistencies, and geocoding case locations to allow for spatial analysis. Since real-time data can be noisy and incomplete, sophisticated algorithms will be used to fill in missing data points and standardize the information for further analysis.

- Spatial Analysis

Geographic Information Systems (GIS) will be used to visualize the geographical distribution of cases and identify spatial clusters where transmission is particularly intense. This involves plotting case locations on a map and overlaying them with environmental data to detect correlations between disease incidence and factors like population density, temperature, and humidity. Techniques such as kernel density estimation will be employed to highlight areas with a high concentration of cases, while GWR and LISA will be used to examine the relationship between socioeconomic factors and disease spread.

- Modeling

The core of the project will be the development of a dynamic epidemiological model that simulates the transmission of the disease in real time. Compartmental models like the Susceptible-Infectious-Recovered (SIR) or Susceptible-Exposed-Infectious-Recovered (SEIR) framework will be used to model the disease's progression. These models will be enhanced with agent-based modeling techniques that simulate individual behaviors and their impact on disease transmission. Real-time data will be fed into these models to ensure they are continuously updated with the latest information.
Graph Attention Networks (GAT) will be employed to capture both spatial and temporal dependencies, enabling the model to predict where and when outbreaks are likely to occur next. This model will be particularly effective for diseases like dengue and Zika, where environmental conditions and human mobility patterns play a critical role in transmission dynamics.

- Visualization

The final step in the methodology is to develop an interactive web-based platform that allows users to visualize the epidemic map in real time. This platform will display the geographical distribution of cases, highlight high-risk areas, and provide model predictions for the future trajectory of the epidemic. The platform will be designed to be user-friendly, allowing public health officials to easily access and interpret the data. Features such as zooming, filtering by timeand location will be included to ensure flexibility and usability for various stakeholders. Public health authorities can use this platform to quickly assess the current situation, identify hotspots, and prioritize interventions accordingly.

5. ***OUTCOMES***

The expected outcomes of the project include:

- A Real-Time Mapping System: A fully operational, real-time mapping system that visualizes the spread of the epidemic, offering up-to-date data on the distribution of cases and predicted outbreak trajectories.
- Identification of High-Risk Areas: The system will highlight regions where disease transmission is likely to escalate, allowing public health officials to allocate resources effectively and target interventions in areas with the greatest need.
- Improved Understanding of Disease Transmission: By incorporating spatial and temporal data, the project will provide new insights into the factors that influence disease spread, including the impact of environmental conditions and human mobility patterns.
- Enhanced Public Health Preparedness: The system will help authorities respond more rapidly to

emerging outbreaks, ensuring that interventions are implemented in a timely and efficient manner.
- Scalability: The system will be designed to be flexible and scalable, allowing it to be applied to different diseases and geographical contexts, making it a valuable tool for future epidemics.

6. TIMELINE OF THE PROJECT/ PROJECT EXECUTION PLAN

The project will be executed over a period of six months, broken down into the following phases:
- Month 1-2: Data Collection and Preprocessing
  During this phase, the team will gather epidemiological, environmental, and mobility data from various sources. This data will then be cleaned, standardized, and geocoded for use in the spatial analysis and modeling stages.
- Month 3: Development of Spatial Analysis Tools
  GIS-based tools will be developed to visualize the spread of the disease. This includes building maps to highlight the geographical distribution of cases, as well as spatial analysis techniques like kernel density estimation and GWR.
- Month 3: Model Development and Integration
  The epidemiological models, including the SIR and SEIR frameworks, will be integrated with real-time data sources. The team will also develop the machine learning models, such as GAT, to dynamically predict future outbreaks.
- Month 4: Testing and Validation
  The models will be validated using historical outbreak data and real-time data feeds. The team will evaluate the models' accuracy by comparing predictions with actual outcomes and adjusting the algorithms as needed.
- Month 4: Visualization Platform Development and Final Report
  The interactive platform for real-time mapping will be completed. The team will also compile the project's findings and prepare a final report, summarizing the system's performance and offering recommendations for future improvements.

7. CONCLUSION

Real-time mapping of epidemic spread is an essential tool in the modern fight against infectious diseases. By integrating epidemiological data with spatial analysis and advanced modeling techniques, health authorities can gain a clearer understanding of how diseases spread and implement timely, targeted interventions. The proposed system will contribute to improved public health management by offering real-time insights into epidemic dynamics and guiding efficient resource allocation. This proactive approach will not only enhance the ability to control current outbreaks but also improve preparedness for future epidemics.
Real-time mapping of epidemic spread is a vital component of modern public health efforts. By leveraging advanced technologies and data analytics, stakeholders can enhance their ability to monitor, respond to, and ultimately mitigate the impact of infectious diseases. As the world continues to face emerging health threats, the importance of effective mapping and response strategies will only grow, making this domain a critical area for research and innovation.

REFERENCES

1. Agudelo, S., & Ventresca, M. "Modeling the spread of the Zika virus by sexual and mosquito transmission." *PLOS ONE* (2022).
2. Zhu, X., Zhang, Y., Ying, H., Chi, H., Sun, G., & Zeng, L. "Modeling epidemic dynamics using Graph Attention-based Spatial Temporal networks." *PLOS ONE* (2024).
3. Sulistyawati, S. "Measuring the dengue risk area using Geographic Information System." *Insights in Public Health Journal* (2020).
4. Liu, Q., et al. "Real-time forecasting of influenza outbreaks using digital surveillance systems." *Journal of Medical Internet Research* (2019).
5. Petropoulos, F., & Chhabra, S. "Short-term epidemic trend predictions using statistical models."

BMC Public Health (2020).
6. Liu, H., et al. "A comparative study on influenza outbreak predictions using LSTM and Random Forest models." *Computational and Mathematical Methods in Medicine* (2019).
7. Nikparvar, B., & Giuliani, M. "Spatio-temporal models for epidemic forecasting using mobility data." *Journal of Medical Internet Research* (2021).
8. Wang, H., et al. "Machine learning models for epidemic forecasting: COVID-19 in Germany." *BMC Infectious Diseases* (2020).
9. Da Silva, T., & Gao, J. "Real-time spatio-temporal analysis for disease forecasting: A machine learning approach." *PLOS Computational Biology* (2021).
10. Kane, M., & Zhu, Y. "Advances in epidemic forecasting: Deep learning approaches for avian influenza." *Infectious Disease Modelling* (2018).
11. Ginsberg, J., et al. "Detecting influenza epidemics using search engine query data." *Nature* (2009).
12. Li, R., et al. "Forecasting influenza activity in the United States using machine learning techniques." *PLOS Computational Biology* (2017).
13. Yang, W., et al. "Real-time epidemic forecasting for pandemic influenza." *Proceedings of the National Academy of Sciences* (2015).
14. Shaman, J., & Karspeck, A. "Forecasting seasonal outbreaks of influenza." *Proceedings of the National Academy of Sciences* (2012).
15. Reich, N., et al. "A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States." *Proceedings of the National Academy of Sciences* (2019).
16. Viboud, C., et al. "The RAPIDD Ebola forecasting challenge: Synthesis and lessons learnt." *Epidemics* (2018).
17. Funk, S., et al. "Assessing the performance of real-time epidemic forecasts: A case study of the 2013–2016 West African Ebola outbreak." *PLOS Computational Biology* (2019).
18. Rivers, C., et al. "Modeling the impact of interventions on an epidemic outbreak in real time: The 2014 Ebola case." *PLOS Currents Outbreaks* (2014).
19. Cauchemez, S., et al. "Real-time estimates in early stages of the 2009 influenza A/H1N1 pandemic." *PLOS ONE* (2009).
20. Viboud, C., & Vespignani, A. "The future of influenza forecasts." *Proceedings of the National Academy of Sciences* (2019).