© November 2025 | IJIRT | Volume 12 Issue 6 | ISSN: 2349-6002

OutbreakAtlas: Advanced Epidemic Alert System for Instant Epidemic Monitoring and Hotspot Identification

Ravindranath A[1], Rohit Rajagopal[2], Shubhankar Vishwakarma[3], Uday Sankar[4], Dr A Prakash[5]
[1,2,3,4]*Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru,Karnataka*
[5]*Professor, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru Karnataka*

***Abstract*—For an effective resolution of an infectious disease outbreak, timely detection and intervention are essential. OutbreakAtlas serves as a proactive alert system for epidemics. Using a density-based spatial clustering technique, it will almost instantly identify disease hotspots and produce map visualizations of them. The density-based spatial clustering (DBSCAN) algorithm will be used by the system to gather geotagged reported incidence cases from online sources. This algorithm quickly displays filtered data on an interactive map by classifying and identifying geo-tagged reported cases. Any reports pertaining to a high-risk hotspot mentioned in the original report are highlighted. An important improvement over traditional epidemiological surveillance systems is the integration of machine learning spatial clustering algorithm, crowdsourced case reporting, and interactive mapping. Additionally, the detection and tracking of potential epidemic threats will be accelerated by combining spatial clustering techniques with near real-time machine learning across large regions and dense epidemiological data. This will improve centralized reporting of emerging risks at the local and regional levels.**

***Index Terms*—Epidemic Surveillance, DBSCAN, Clustering, NLP,Public Health Informatics, Real-Time Visualization, Outbreak Detection.**

I. INTRODUCTION

Global health, economic stability, and social well-being have all been seriously threatened by infectious diseases. The significance of early detection and coordinated responses is underscored by recent outbreaks such as COVID-19, Ebola, Zika, SARS, MERS, H1N1 influenza, cholera, and dengue fever. There are several ways that infectious diseases can spread. Compared to viruses spread by vectors or tainted water, airborne viruses are more easily spread. On the other hand, any infectious disease can spread quickly if early warning systems are insufficient. Conventional monitoring techniques depend on laboratory testing to validate cases; however, the way incidents and case reports are recorded frequently causes reporting delays. Days or even weeks may pass before authorities are alerted to a case.

Existing frameworks for monitoring epidemics, including WHO-EWARN, CDC's FluView, and national health information systems, rely primarily on case reports from hospitals and public health agencies. Even though these systems can effectively monitor conditions within regulated health environments, they rely heavily on hospitals and public health organizations for updates on infectious diseases.It can be difficult to get up-to-date information and data in rural or underserved areas, where outbreaks frequently occur.Real-time monitoring and the reporting of infectious disease responses are hampered by bureaucratic processes that assign tasks, which leads to an over-reliance on centralized reporting systems. Data inconsistencies and fragmented information flows make it difficult to identify emerging pathogen hotspots in a timely manner. Infectious diseases like dengue and malaria are affected by changes in the seasons and climate.

In response to these challenges, Outbreak Atlas was developed as an epidemic intelligence platform. It combines interactive geospatial visualization, machine learning-based clustering, and crowdsourced data collection. Through this platform, people and organizations that provide healthcare can work together to report and visualize new outbreaks in real time. Outbreak Atlas finds and highlights clusters of reported cases using the DBSCAN (Density-Based

Spatial Clustering of Applications with Noise) algorithm. This makes it possible to distinguish real hotspots from sporadic variations. As a result, public health organizations can take proactive measures by allocating funds, starting containment initiatives, and releasing regional alerts. Outbreak Atlas is a scalable, data-driven framework that combines computational intelligence with community-driven public health initiatives to reduce the spread of diseases in a variety of contexts, from isolated cholera outbreaks and large respiratory infections to urban dengue clusters.

II. LITERATURE SURVEY

Machine learning methods that improve surveillance and predict disease outbreaks are being studied by Ekundayo (2024) [1] In addition to other unsupervised learning techniques, the study uses supervised techniques like Random Forest, SVM, and DBSCAN to integrate genomic data, social media activity, climate data, and electronic health records.The results show that ML-driven systems outperform conventional models in terms of speed and accuracy.

TepiSense, an AI platform created for real-time epidemic monitoring, was introduced by Tahir and Mehmood (2025) [2] Their system uses deep learning models like CNNs, BLSTMs, mBERTs, and mT5 to remove irrelevant data from social media They created the EpiBank database, which contains 271 million tweets about six different illnesses.

Aryffin et al. (2025) [3] carried out a systematic review of 69 studies published between 2019 and 2023 to determine technological trends for epidemic intelligence.

Al-Ahmadi and Al-Amri (2022) [4] showed how well the DBSCAN algorithm works with spatial clustering to identify COVID-19 hotspots. Their findings demonstrated that density-based algorithms perform better than more traditional methods like K-Means when it comes to managing outliers and non-standard cluster formations. This improves the OutbreakAtlas system and confirms DBSCAN as a trustworthy option for locating epidemic hotspots.

According to Cordes et al. (2017) [5], the WHO developed the Early Warning Alert and Response Network (EWARN) to detect epidemics quickly. This concept has been incorporated by OutbreakAtlas to allow for real-time alerts.

Badker et al. (2025) [6] suggested a global open-access epidemic database that compiles surveillance reports from 237 nations, covering over 170 pathogens.This dataset sets a benchmark for upcoming epidemic forecasting models and highlights the importance of open and compatible data-sharing systems in global health.

A review of clustering techniques for syndromic surveillance systems was carried out by Yeng, Woldaregay, and Solvoll (2020) [7]. DBSCAN and scan statistics are better at detecting disease clusters with irregular spatial configurations, according to the study, which focused on spatial, temporal, and spatio-temporal techniques. This supports OutbreakAtlas' clustering methodology.

The difficulties of real-time epidemic forecasting, in particular the need for timely, high-quality data, were covered by Desai et al. (2019) [8]. In order to improve forecasting accuracy, the authors emphasized the significance of including population dynamics and mobility data. OutbreakAtlas's focus on continuous live data collection is influenced by these observations.

Insights from the COVID-19 pandemic were examined by Rosenfeld, Tibshirani, and Guntuboyina (2021) [9]. To effectively track epidemics, they emphasized the necessity of digital data workflows and federated surveillance systems. Their study supports the case for decentralized systems, such as OutbreakAtlas, which make it easier to collect and visualize epidemic data.

Using spatiotemporal data, Wadhwa, Bhardwaj, and Verma (2022) [10] suggested a rapid surveillance strategy for COVID-19. By using real-time geographic mapping, they were able to identify hotspots early.

According to their findings, combining spatial and temporal characteristics greatly improves the sensitivity and promptness of epidemic alert systems—a critical feature that OutbreakAtlas also reflects.

III. METHODOLOGY

The OutbreakAtlas framework integrates contemporary web technologies, geospatial analysis, and machine learning to deliver immediate epidemic information. This system is designed with a modular and layered approach. Such a configuration facilitates expansion, rapid data processing, and swift reactions to different epidemic scenarios. It consists of six primary components: data collection, preprocessing, clustering, visualization, alerting, and configuration management.

3.1 System Architecture Overview

The following layers comprise the four-tier architecture of the OutbreakAtlas system:

1.The Frontend Layer i.e User Interface collects geotagged epidemic reports and displays real-time hotspot visualizations.

2. Data ingestion, validation, and storage are handled by the Backend Layer (Data Processing and API Engine).

3. The Analytics Layer (Machine Learning Core) performs hotspot analysis and spatial clustering using DBSCAN.

4. The Visualization and Alert Layer creates interactive maps and sends alerts to users.

To enable the smooth transfer of data from user inputs to public dashboards, these layers communicate with one another through RESTful APIs. The architecture supports many users and real-time updates while ensuring low latency and quick responses.

3.2 Information Collection and User Interface

OutbreakAtlas gathers data from public health data feeds, automated web scrapers, and user input. A simple online form that includes the following allows users to report outbreaks:

Location (longitude and latitude)

The disease's name and a list of its symptoms

The report's date and time.

He number of cases or severity level

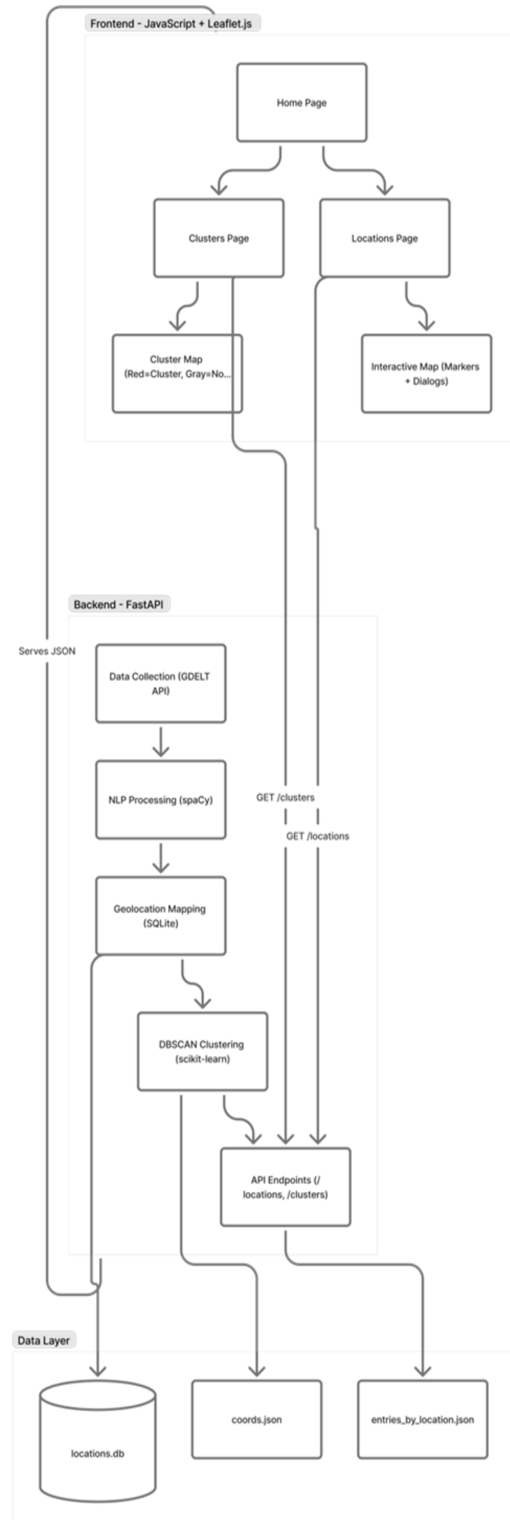

FIG.1 Architecture of OutbreakAtlas

Here HTML, CSS, and JavaScript are used to create the frontend interface, and Leaflet.js is used for mapping. Simplicity, accessibility are given priority in the design to encourage involvement from the general public, medical professionals, and public health officials. Fast API is used to send reports to the backend through REST APIs.

3.3 Preparing and Verifying Data

Before being entered into the database, received reports go through a validation process. These crucial preprocessing tasks are finished by the Python backend, which was developed using FastAPI:

Repetitive entries are removed through duplicate removal.

-Data cleaning fixes fields that are inconsistent or missing.

-Valid latitude and longitude coordinates are guaranteed by geospatial verification.

-Only recent reports are kept by Temporal Filtering according to predetermined time periods (such as the previous 30 days).

Different data formats are standardized into a unified structure through normalization.

SQLAlchemy ORM is used to store the cleaned data, which is linked to PostgreSQL/MySQL for production and SQLite for local development. For effective searching, each entry is arranged spatially and assigned a timestamp

3.4 Hotspot Detection and Clustering

OutbreakAtlas is analytically based on the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. For real-world epidemic analysis, DBSCAN was chosen because of its capacity to identify clusters of different shapes and to remove noise from data that is unevenly distributed.

Steps in the Algorithm:

1. Use the Haversine distance formula to calculate the geographic distances between reported cases pairwise.
2. Establish the minimum number of cases (MinPts) and neighborhood radius (Eps) needed to form a cluster.
3. Find dense regions that meet these requirements, then group them into clusters.
4. Isolated cases should be treated as noise or low-risk indicators.
5. To maintain cluster stability over time, apply temporal smoothing.

Technology stack:

a. Scikit-learn:provides DBSCAN implementation.
b. NumPy and Pandas: Utilized for data manipulation and feature extraction.
c. Shapely & GeoPandas: permits spatial functions such as indexing, union, and buffering.
d. Matplotlib/GeoJSON: aids in cluster rendering and map integration.

Each cluster is assigned a severity score based on density, duration, and recent activity. Clusters exceeding preset severity thresholds trigger real-time alerts.

3.5 Dashboard and Geospatial Visualization

Using Leaflet.js, the visualization module creates interactive geospatial maps from clustered data. Both regional and global viewpoints on outbreak dynamics are provided by the dashboard.

Important Visualization Elements:

1. Identified clusters are shown as red circles on the Real-Time Hotspot Map, with sizes based on case density.
2. Maps automatically update with new reports thanks to dynamic updates.
3. By clicking on a hotspot, users can access a comprehensive panel that provides an overview of the local outbreak through Cluster Inspection.
4. For the best rendering, GeoJSON Conversion reformats cluster data as GeoJSON.
5. Red (critical), orange (emerging), and green (stable) statuses are indicated by color-coded alerts.

This interface connects user interaction with analytical findings. It ensures that the situation is promptly communicated to the public and health authorities.

Fig. 2. Global visualization of identified epidemic hotspots using DBSCAN clustering in OutbreakAtlas.

3.6 Information and Alert Panel for the Region

As shown in Fig. 3, OutbreakAtlas displays an interactive regional alert panel that summarizes ongoing outbreaks when a user clicks on a hotspot.

a. The system retrieves real-time news and outbreak feeds for the designated area, providing information such as the name of the disease and the affected area.
b. An overview of the outbreak's causes
c. The number of cases reported and the pertinent date

For instance, choosing India displays warnings for H3N2 influenza, norovirus, listeria, and diarrhea outbreaks. Users and health organizations can quickly assess the local epidemic scenario with the help of this information.

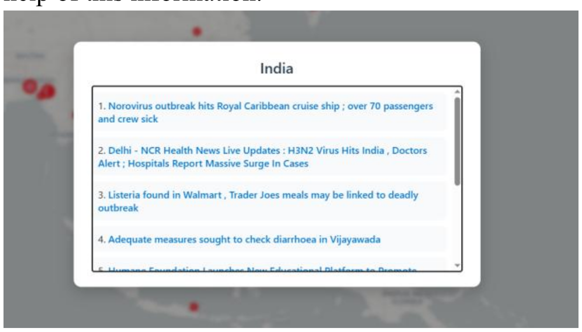

Fig. 3. Regional outbreak alert interface displaying real-time epidemic reports for India.

III. RESULTS AND DISCUSSION

The DBSCAN clustering algorithm was compared to other popular techniques, such as K-Means, Hierarchical Clustering, and Gaussian Mixture Models (GMM), in order to assess OutbreakAtlas's performance. DBSCAN performed better than the other algorithms in important areas like cluster shape flexibility, noise management, spatial fitting, and scalability, as shown in Fig. 4. DBSCAN effectively identified irregularly shaped outbreak clusters and controlled for spatial noise, which is commonly present in real-world epidemic data, in contrast to centroid-based approaches that presume clusters will have circular shapes. Additionally, its ability to identify outliers improved hotspot identification's dependability, especially in a variety of reporting contexts.

Clustering quality was quantitatively evaluated using two internal assessment metrics: the Calinski-Harabasz Index (CHI) and the Davies-Bouldin Index (DBI) The DBI score of 1.05124 indicated that the clusters were compact and well-separated. On the other hand, the clusters' internal cohesion varied significantly, as indicated by the CHI score of 26.92051. According to these metrics, the DBSCAN-based model produces spatially coherent and consistent clusters that are useful for tracking and visualizing epidemics in real time. However, conventional methods such as K-Means and Hierarchical algorithms were less effective for large, unstructured epidemic datasets because they generated fewer distinct clusters and were more prone to noise.

In summary, the findings suggest that DBSCAN is the optimal choice for the OutbreakAtlas framework. It strikes a favorable balance between accuracy, computational efficiency, and flexibility. Its high scalability and capability to manage dynamic, crowd-sourced geospatial data render it especially appropriate for monitoring a range of epidemic scenarios, from small disease clusters to worldwide outbreak surveillance. By employing this clustering technique, OutbreakAtlas can deliver dependable, noise-resistant, and actionable epidemic intelligence across various data densities and geographic regions.

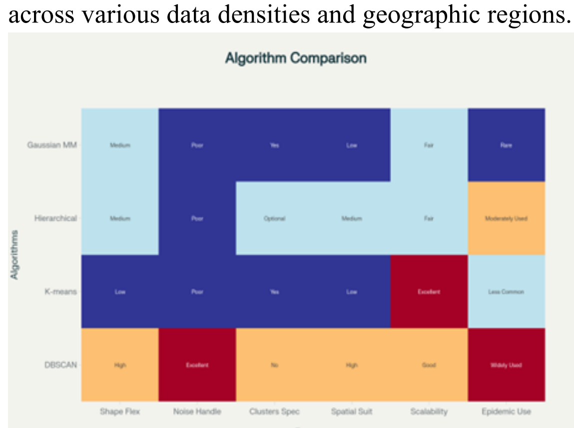

Fig. 4. Algorithm Comparison of DBSCAN, K-Means, Hierarchical, and Gaussian Mixture Models based on key clustering performance features and epidemic suitability

V. CONCLUSION

OutbreakAtlas, an epidemic intelligence platform that combines geospatial analysis, data science, and public health, is presented in this study. It provides outbreak visualization and real-time detection. For density-based spatial analysis, the platform uses the DBSCAN clustering algorithm. It distinguishes real epidemic signals from noise and identifies changing disease hotspots. Rapid processing, smooth data flow, and interactive visualization of disease trends at both the global and regional levels are made possible by the combination of FastAPI, Leaflet.js, and GeoPandas. The platform can produce spatially coherent, high-quality clusters with little delay, according to experimental results confirmed by strong internal validation metrics (Calinski-Harabasz Index: 26.92051, Davies-Bouldin Index: 1.05124). Rapid, data-driven responses to infectious disease outbreaks are made possible by this capability.

In addition to being technologically advanced, OutbreakAtlas signifies a move toward participatory and AI-enhanced public health surveillance. Its architecture encourages transparency, scalability, and accessibility. Collaboration between the general public, research institutions, and health authorities is promoted by this framework. Because the system can adapt to various epidemiological scenarios, it can be used to monitor a wide range of diseases, including COVID-19, cholera, dengue, influenza, and Ebola. OutbreakAtlas may eventually incorporate predictive forecasting models, federated data integration, and Internet of Things-enabled health monitoring. An intelligent ecosystem would be produced as a result, enhancing global readiness and epidemic resistance.

VII. FUTURE WORK

Pandemics are identified and shown in real time by OutbreakAtlas. However, a few changes could increase its efficacy and make scaling easier: 1 IoT and Sensor Integration: Set up health monitoring systems that make use of wearable sensors, the Internet of Things, and environmental data. To help with early problem detection and outbreak prediction, this includes temperature, humidity, and air quality data.

2. Edge and Cloud-Native Deployment:
Convert the system to a microservices architecture that is cloud-native and integrates edge computing. During widespread health emergencies, this will allow for faster processing, better scalability, and lower latencies.

3. Developing Mobile Apps: Release a mobile application to broaden OutbreakAtlas' audience. To improve accessibility, it will provide real-time alerts, user-submitted health updates, and GPS-based local warnings.

4. Multilingual and Inclusive Interface:
Offer accessibility features and multilingual support. As a result, many people will be encouraged to participate, including those who are not very tech-savvy.

5. Misinformation Detection and Data Validation:
Methodically remove inaccurate or misleading information about epidemics from public data sources like social media by using natural language processing techniques.

6. To promote cooperation while preserving privacy, use federated learning frameworks. They will make it possible for laboratories, medical facilities, and public health organizations to work together to train models while protecting private or sensitive information.

REFRENCES

[1] F. Ekundayo, "Using Machine Learning to Predict Disease Outbreaks and Enhance Public Health Surveillance," World Journal of Advanced Research and Reviews, vol. 24, no. 3, pp. 794–811, 2024.

[2] B. Tahir and M. A. Mehmood, "TepiSense: A Social Computing-Based Real-Time Epidemic Surveillance System Using Artificial Intelligence," IEEE Access, vol. 13, pp. 25180–25210, 2025.

[3] H. A. K. Aryffin, M. A. B. Sahbudin, S. Ali, and A. O. Khadidos, "Technological Trends in Epidemic Intelligence for Infectious Disease Surveillance: A Systematic Literature Review," BMC Public Health, vol. 25, no. 1, 2025.

[4] S. Al-Ahmadi and A. Al-Amri, "Data Clustering Mapping of COVID-19 Pandemic Based on DBSCAN Algorithm," International Journal of Computer Science and Network Security (IJCSNS), vol. 22, no. 4, pp. 1–8, 2022.

[5] K. M. Cordes, S. T. Cookson, A. T. Boyd, C. Hardy, M. R. Malik, P. Mala, and S. M. Perera, "Real-Time Surveillance in Emergencies Using the Early Warning Alert and Response Network (EWARN)," Emerging Infectious Diseases, vol. 23, no. 13 (Suppl), pp. S131–S137, 2017.
[6] R. Badker, A. Morrison, W. J. Alonso, S. Chakraborty, A. Cori, C. A. Donnelly, and V. Colizza, "Constructing a Global Human Epidemic Database Using Open-Source Surveillance Reports," Scientific Data, vol. 12, Article 145, 2025.
[7] P. K. Yeng, A. Z. Woldaregay, and T. Solvoll, "Cluster Detection Mechanisms for Syndromic Surveillance Systems: Systematic Review and Framework Development," JMIR Public Health and Surveillance, vol. 6, no. 4, e21512, 2020.
[8] A. N. Desai, M. U. G. Kraemer, S. Bhatia, A. Cori, P. Nouvellet, M. Herringer, and J. S. Brownstein, "Real-Time Epidemic Forecasting: Challenges and Opportunities," Health Security, vol. 17, no. 4, pp. 268–275, 2019.
[9] R. Rosenfeld, R. J. Tibshirani, and A. Guntuboyina, "Epidemic Tracking and Forecasting: Lessons Learned from a Tumultuous Year," Proceedings of the National Academy of Sciences (PNAS), vol. 118, no. 51, e2111456118, 2021.
[10] A. Wadhwa, A. Bhardwaj, and V. S. Verma, "Rapid Surveillance of COVID-19 by Timely Detection of Emerging Hotspots Using Spatio-Temporal Data," Journal of Healthcare Engineering, Article ID 2583187, 2022.