GSC Advanced Research and Reviews

eISSN: 2582-4597 CODEN (USA): GARRC2

Cross Ref DOI: 10.30574/gscarr

Journal homepage: https://gsconlinepress.com/journals/gscarr/

GSC Advanced Research and Reviews

GSC Online Press INDIA

(REVIEW ARTICLE)

Check for updates

Data visualization in diseases epidemiology

Adeleye Adewuyi [1, *], Victor Adeoye [2], Deborah Ave Okon [3], Emmanuel Faderin [4], Olumayowa Adeleke Idowu [5] and Taiwo Akindahunsi [6]

[1] Department of Computer Science, University of People, Pasadena, USA.
[2] InStrat Global Health Solutions Ltd., Nigeria.
[3] Department of Public Health, New Mexico State University, Las Cruces, USA.
[4] Federal Neuro-Psychiatric Hospital, Yaba Lagos, Nigeria.
[5] Department of Economics, University of Pittsburgh, Pittsburgh, USA.
[6] Lagos State Primary Health Care Board, Yaba Lagos, Nigeria.

GSC Advanced Research and Reviews, 2021, 09(03), 151–163

Publication history: Received on 08 September 2021; revised on 25 October 2021; accepted on 28 October 2021

Article DOI: https://doi.org/10.30574/gscarr.2021.9.3.0231

Abstract

Data visualization has transformed diseases epidemiology by empowering researchers, practitioners, and policymakers to unleash complex patterns, trends, and relationships. It uncovers hidden correlations and clusters, tracks disease outbreaks and transmission dynamics, identifies high-risk populations and areas, evaluates intervention effectiveness, and communicates complex findings to diverse audiences. Through exemplary visualizations, data visualization distills complex epidemiological data into actionable insights, informing data-driven decisions that promote public health and well-being. As advanced visualization techniques continue to evolve, they accelerate the understanding of disease dynamics, aid in the allocation of resources, and drive proactive strategies for disease mitigation. However, barriers such as data quality, infrastructure limitations, and the need for skilled personnel persist, especially in under-resourced settings. This paper presents a critical evaluation of the role of data visualization in epidemiological practice, discussing its implications for risk identification, policy formulation, and the proactive management of health crises. The insights gained from this paper will illuminate pathways for future innovations in disease surveillance and control.

Keywords: Data Visualization; Epidemiology; Disease Dynamics; Public Health

1. Introduction

The term 'data visualization' has a long history, dating back to the 2nd century AD [1]. In ancient societies, drawings and other visual representations were used to explore the world and document historical events [2]. However, the invention of computer technology has dramatically transformed how data is visually represented. With computer-based graphical data visualization, data analysis has become faster and more accurate [3]. Today, data visualization is integral to research across various fields, including public health, algorithms, human perception, animation, and computer vision. In diseases epidemiology, it involves translating complex disease-related data into easily comprehensible visual formats [4]. These formats—such as charts, graphs, maps, and dashboards—enable researchers, policymakers, and the public to identify patterns, trends, and relationships in large datasets. Given the vast datasets used in epidemiological research, which often cover multiple variables across time and geographic locations, data visualization simplifies this complexity, offering numerous advantages. It helps identify disease patterns and trends, enabling the detection of outbreaks, changes in disease incidence, and geographic hotspots [5]. Moreover, it enhances communication, ensuring that data is accessible not only to experts but also to a wider audience, including the public and decision-makers. This capacity to distill large datasets into meaningful insights makes data visualization an indispensable tool for effective

* Corresponding author: Adeleye Adewuyi.

decision-making in public health, particularly when real-time responses are necessary, such as during infectious disease outbreaks [6-7].

The use of data visualization in epidemiology has a rich history, dating back to the mid-19th century. One of the most iconic early examples is the work of John Snow during the 1854 cholera outbreak in London [8]. Snow used spatial mapping to track cholera deaths and identified a cluster of cases around a specific water pump. His visual representation of the data led him to hypothesize that cholera was waterborne, fundamentally altering public health approaches to disease control [9]. This historic example highlighted the potential of visualizing epidemiological data to reveal underlying disease patterns. Since Snow's time, data visualization in epidemiology has evolved, moving from simple, static maps to sophisticated dynamic tools that include interactive dashboards, heatmaps, and geographic information systems (GIS). These modern tools provide real-time insights, allowing health authorities to monitor disease outbreaks more efficiently and respond more swiftly.

The role of data visualization in public health decision-making cannot be overstated. During the COVID-19 pandemic, real-time dashboards, such as the Johns Hopkins University COVID-19 dashboard, became central to global monitoring efforts. By visually displaying case counts, deaths, and vaccination progress, these tools enabled governments and organizations to make informed decisions about public health interventions and resource distribution [10]. Furthermore, well-designed visualizations are essential to risk communication, helping policymakers and the public grasp health risks and understand the importance of preventive measures. In this way, data visualization supports not only early detection and monitoring but also the formulation of evidence-based policies that target high-risk areas and allocate resources effectively [11-12].

However, several challenges complicate the use of data visualization in epidemiology. One of the most significant issues is the quality and completeness of data. In many low- and middle-income countries, underdeveloped health surveillance systems produce incomplete datasets, making it difficult to generate accurate visualizations [13]. Moreover, there is ethical concern, particularly in relation to privacy and data security. Epidemiological data often include sensitive information, and visualizations that are too detailed can breach individuals' privacy [14]. Hence, balancing the need for transparency with ethical obligations remains a challenge for data visualization in epidemiology.

This paper aims to explore the role that data visualization plays in epidemiology, focusing on its support for public health decision-making. Additionally, this paper will assess the effectiveness of current visualization techniques in conveying complex epidemiological data and identify challenges and limitations, including issues of data quality, ethical concerns, and the risk of misinterpretation. Furthermore, it will propose potential solutions and future trends that can enhance the utility of data visualization in disease surveillance and public health interventions.

2. Types of Data Visualization Techniques in Epidemiology

In epidemiology, various data visualization techniques are employed to help public health professionals and researchers effectively communicate patterns, trends, and insights derived from complex datasets. Each technique serves a distinct purpose, whether for descriptive analysis, spatial mapping, network representation, or temporal tracking. Below are the primary data visualization techniques used in epidemiology.

2.1. Descriptive Visualizations: Line Charts, Bar Charts, and Histograms

Descriptive visualizations are the most widely used forms of data representation in epidemiology. These types of visuals are useful for summarizing metrics such as incidence, prevalence, and mortality rates, thus, providing a snapshot of the epidemiological situation [15]. Their primary function is to simplify complex information and allowing researchers, public health officials, and the general public to quickly grasp health trends and patterns. This approach not only aids in understanding but also enhances the effectiveness of communication in public health, making it easier to convey information.

Several types of descriptive visualizations commonly used in epidemiology, each serving distinct purposes includes:

- Line Charts: Line charts are used to represent changes in disease metrics (such as case counts, hospitalizations, or deaths) over time. Each point on the chart represents a data point for a specific time interval, connected by a line that shows trends, fluctuations, and seasonality. In epidemiology, line charts are used to depict the number of new cases (incidence) or cumulative case counts over days, months, or years [16-17]. For example, during the COVID-19 pandemic, line charts were frequently used to illustrate the rising number of cases or deaths over time, allowing health authorities to identify peaks in transmission and assess the impact of interventions [18].

- Bar Charts: Bar charts are used for comparing different groups or categories in a dataset. They display data using rectangular bars, where the length of each bar is proportional to the value it represents [19]. In epidemiology, bar charts are used to compare disease prevalence or incidence rates across different population groups, such as age, gender, or geographic location. They are also useful for showing categorical data, such as the distribution of different diseases within a population or comparing vaccination rates across regions [15, 19].
- Histograms: Histograms are a specific type of bar chart that represents the distribution of continuous data [20]. In epidemiology, histograms are useful for displaying the frequency distribution of variables such as age at diagnosis, duration of illness, or time between infection and onset of symptoms. Histograms are often used to describe the spread and skewness of epidemiological data, helping researchers and public health professionals identify patterns such as early onset or the clustering of cases around certain variables [15, 21].

2.2. Spatial Visualizations: Geographic Information Systems (GIS), Heatmaps, and Choropleth Maps

Spatial visualizations are essential for mapping disease spread, identifying hotspots, and visualizing the geographic distribution of diseases. These tools allow epidemiologists to analyze how diseases vary across regions and to understand environmental or geographic factors contributing to disease transmission [22]. Several spatial visualization tools include:

- Geographic Information Systems (GIS): GIS is a powerful tool in epidemiology that integrates location-based data with epidemiological information to create detailed maps that visualize disease patterns [23]. By layering various datasets—such as population density, climate, healthcare access, and disease occurrence—GIS allows researchers to analyze how geographic and environmental factors influence the spread of diseases [24]. For example, GIS has been used extensively in malaria research to map regions where conditions are favorable for mosquito breeding and to identify areas requiring vector control interventions. GIS mapping is also used for tracking the spread of diseases like cholera, where water sources and sanitation facilities are key factors in transmission [25].
- Heatmaps: Heatmaps represent data using a color gradient to show the intensity of a variable across different regions [26]. In epidemiology, heatmaps are often used to identify geographic hotspots where disease incidence is high. Darker colors indicate regions with higher case counts, while lighter colors represent areas with fewer cases [27]. Heatmaps are particularly useful for visualizing infectious disease outbreaks, as they allow public health authorities to target interventions such as vaccination campaigns or quarantine measures in high-incidence areas. They have been used during the COVID-19 pandemic to track the intensity of virus transmission in different countries or regions [28-29].
- Choropleth Maps: Choropleth maps are a type of spatial visualization where geographic areas (e.g., countries, states, or districts) are shaded or colored based on a statistical variable, such as disease incidence or vaccination coverage [30]. Each geographic unit is filled with a color that corresponds to the value of the variable being represented. Choropleth maps are frequently used in epidemiology to illustrate disease prevalence or mortality rates across different regions [31]. For example, a choropleth map of HIV prevalence might show higher rates in sub-Saharan Africa compared to other regions, highlighting the areas where prevention and treatment efforts should be focused [32].

2.3. Network Visualizations: Representing Transmission Patterns and Contact Tracing

Network visualizations are used in epidemiology to represent relationships and interactions between individuals or entities, particularly in disease transmission and contact tracing [33-34].

- Transmission Networks: Transmission network visualizations depict how a disease is transmitted between individuals or groups. Each node in the network represents an individual, while the edges (connections between nodes) represent interactions or contacts that may lead to transmission [35]. Network visualizations have been particularly useful in studying diseases that spread through direct contact, such as sexually transmitted infections (STIs) or respiratory diseases. For example, during the early stages of the COVID-19 pandemic, network visualizations were used to map the spread of the virus within families, workplaces, and communities, allowing public health officials to identify 'super spreaders' or events that significantly contributed to transmission [36-37].
- Contact Tracing: Contact tracing visualizations help epidemiologists track the spread of diseases by visualizing the relationships between infected individuals and their contacts [37]. These visualizations are essential for identifying people who may have been exposed to a disease and for preventing further transmission. During

outbreaks of diseases like Ebola and COVID-19, contact tracing networks helped health authorities follow the path of transmission and isolate potentially infected individuals before they spread the disease to others.

2.4. Temporal Visualizations: Time Series Analysis for Tracking Disease Spread Over Time

Temporal visualizations are essential for tracking how diseases evolve over time, capturing trends in disease incidence, prevalence, and mortality across various time intervals [23]. These visualizations help epidemiologists identify seasonal patterns, assess the impact of interventions, and predict future disease trajectories.

- Time Series Analysis: Time series visualizations represent data points across successive time intervals, enabling researchers to analyze trends, cycles, and anomalies [38]. In epidemiology, time series charts are commonly used to track the number of cases, deaths, or hospitalizations over days, weeks, months, or years. These charts can be used to analyze the effects of interventions such as vaccination campaigns or lockdowns, by comparing trends before and after these interventions were implemented [38-40]. For example, during the influenza season, time series charts are used to track weekly flu cases, allowing health authorities to detect peak transmission periods and deploy resources accordingly [41].
- Cumulative Curves: Cumulative curves are another form of temporal visualization that show the total number of cases or deaths over time [42]. Unlike time series charts, which display the change from one period to the next, cumulative curves continuously increase as cases or deaths are added to the total. Cumulative curves are particularly useful for illustrating the overall burden of disease, such as in visualizations showing the total number of COVID-19 cases or deaths during the course of the pandemic [43].
- Seasonality Plots: In epidemiology, seasonality plots show the cyclical nature of disease transmission by visualizing peaks and troughs over time. These visualizations help public health officials anticipate future outbreaks and plan preventive measures accordingly [44]. For example, in the case of vector-borne diseases like dengue or malaria, seasonality plots can highlight periods of increased transmission related to weather patterns, allowing for timely vector control interventions [45].

2.5. Interactive Dashboards: Use of Tools Like Tableau and Power BI in Real-Time Epidemiological Reporting

Interactive dashboards have become an essential tool in modern epidemiology, particularly during disease outbreaks that require real-time monitoring and reporting. These dashboards allow users to interact with various data visualizations, filtering and exploring different aspects of the data in real-time [46].

- Tableau: Tableau is a popular data visualization tool that enables the creation of interactive dashboards and reports [47]. In epidemiology, Tableau is used to combine multiple visualizations—such as maps, bar charts, and time series graphs—into a single, cohesive interface that can be explored by users [48]. During the COVID-19 pandemic, public health agencies used Tableau to create dashboards that allowed users to track case counts, hospitalizations, testing rates, and vaccination progress by region, age group, or other demographic factors [49].
- Power BI: Microsoft's Power BI is another widely used tool for creating interactive data visualizations and dashboards. In epidemiology, Power BI is used to integrate data from multiple sources (such as electronic health records or disease surveillance systems) and display it in real-time. Public health officials can use Power BI dashboards to monitor the status of an outbreak and respond quickly to emerging trends [50]. For example, during disease outbreaks like Ebola, Power BI dashboards helped visualize the geographic spread of the disease, allowing for timely responses such as travel restrictions or the deployment of healthcare resources.

3. Data Sources for Epidemiological Visualization

Accurate and timely data is the foundation of effective epidemiological visualization [23]. The sources of data that feed into these visualizations come from a range of sectors, including health institutions, environmental monitoring agencies, and even mobile technology platforms. These diverse sources enable public health officials to track disease patterns, forecast outbreaks, and develop intervention strategies. This section will explore the data sources for epidemiological visualization, including surveillance data from health agencies, clinical data from hospitals, mobile and social media data, and environmental data for tracking vector-borne diseases.

3.1. Surveillance Data from National and Global Health Agencies

One of the primary sources of data for epidemiological visualization is surveillance data collected by national and global health agencies, such as the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC). These agencies collect, collate, and disseminate vast amounts of health-related data from member countries and other reporting organizations. The data, often gathered through routine surveillance systems, includes vital statistics like disease incidence, prevalence, mortality rates, and vaccination coverage.

- World Health Organization (WHO): The WHO serves as a central repository for global health data, providing access to datasets from member states on a variety of health issues. The WHO's Global Health Observatory (GHO) offers a wealth of data on diseases, vaccination coverage, and health systems, all of which are instrumental in developing visualizations that inform global public health responses [51-52].
- Centers for Disease Control and Prevention (CDC): The CDC, primarily concerned with public health in the United States, collects extensive surveillance data on communicable and non-communicable diseases. It publishes this data through various reports, dashboards, and health data platforms, which are utilized to create visualizations for monitoring disease trends, such as the spread of influenza or foodborne illnesses [53].

3.2. Clinical Data from Hospitals and Laboratories

Clinical data from hospitals and diagnostic laboratories provide another source of information for epidemiological visualization. This data includes patient records, laboratory test results, and treatment outcomes, offering detailed insights into disease characteristics, transmission, and patient demographics. Clinical data allows for more granular visualizations, such as age-specific incidence rates or disease severity trends across different populations [54].

- Electronic Health Records (EHRs): Many hospitals now use electronic health records to store patient data. This allows for large-scale data aggregation and analysis, which can be used to generate visualizations for disease surveillance. For example, in tracking an outbreak, clinical data can show the age distribution of patients, the severity of symptoms, or the efficacy of treatment options [55]. These data are crucial in understanding not only the epidemiology of a disease but also its clinical progression and impact on specific subpopulations.
- Laboratory Data: Laboratories contribute essential data on pathogen identification, resistance patterns, and diagnostic results. For instance, during an outbreak of a novel infectious disease, labs provide critical information on the pathogen's characteristics, which can then be visualized to show the geographic distribution of cases or the evolution of the virus [56]. Laboratory data is particularly important in tracking antimicrobial resistance, where visualizations can illustrate patterns of resistance across different regions or over time [57].

3.3. Mobile and Social Media Data for Real-Time Epidemic Monitoring

Mobile technology and social media platforms have emerged as innovative sources of real-time data for epidemic monitoring and surveillance. This type of data offers a dynamic, immediate picture of how diseases are spreading and how populations are responding to health interventions.

Mobile Data: Mobile phone data, including GPS tracking and call records, is increasingly being used to monitor population movement and behavior during outbreaks [58]. For example, during the COVID-19 pandemic, mobile data was used to assess the effectiveness of lockdown measures by visualizing patterns of human mobility. Similarly, in the case of diseases like cholera, mobile data can track the migration of potentially infected individuals, helping to predict where outbreaks might occur next [59].

Social Media Data: Platforms like Twitter, Facebook, and Google Trends have proven valuable in tracking the public's reaction to health crises and identifying potential hotspots for disease transmission. Social media posts, searches, and interactions are analyzed to detect early signals of an outbreak, such as a sudden spike in discussions about flu symptoms or hospital visits in a specific area [60]. This data can be visualized using heatmaps or word clouds to highlight regions where disease activity may be increasing.

The immediacy and widespread nature of mobile and social media data make them indispensable for real-time epidemic monitoring. This data source helps fill gaps in traditional surveillance systems by providing early warnings about emerging public health threats and enabling swift intervention.

3.4. Environmental and Climate Data to Visualize Vector-Borne Diseases

Environmental and climate data are vital for visualizing vector-borne diseases, which are often influenced by weather patterns, vegetation, and human interaction with the environment. This type of data helps in understanding the ecology of disease vectors, such as mosquitoes or ticks, and predicting the conditions under which these vectors thrive and transmit diseases [61].

Environmental Data: Data on temperature, humidity, rainfall, and land use are used to model the distribution of disease vectors like mosquitoes (in the case of malaria and dengue) or ticks (for Lyme disease) [62]. Visualization tools like Geographic Information Systems (GIS) are used to create maps that show areas at high risk for vector-borne diseases based on these environmental conditions. For example, by combining rainfall data with GIS mapping, health officials can predict the breeding cycles of mosquitoes in tropical regions and target interventions such as insecticide spraying or the distribution of bed nets [61-62].

Climate Data: Climate change is increasingly being recognized as a factor in the spread of vector-borne diseases. Rising temperatures and changing rainfall patterns can expand the habitats of disease vectors, leading to the spread of diseases into new regions [63]. Data visualization can show how climate variables correlate with disease outbreaks, helping to predict future hotspots and guide preventive measures.

4. Challenges in Data Visualization for Epidemiology

While data visualization has proven to be a powerful tool in epidemiology, it is not without challenges. Issues related to data quality, ethical concerns, and visual bias can hinder the effectiveness of visual tools in public health decision-making. This section examines these challenges in detail, addressing how they can compromise the utility and reliability of epidemiological data visualizations.

4.1. Issues with Data Quality, Completeness, and Reliability

One of the primary challenges in data visualization for epidemiology is ensuring data quality, completeness, and reliability. Epidemiological data often comes from diverse sources, such as hospitals, laboratories, national health agencies, and even mobile technologies, all of which may have varying standards for data collection and reporting. Incomplete or inconsistent data can lead to inaccurate visualizations that misrepresent disease trends, potentially resulting in flawed public health interventions [13].

- Inconsistent Reporting: Health data, especially in low-resource settings, may suffer from inconsistent reporting due to infrastructural deficiencies, lack of trained personnel, or outdated health information systems. As a result, epidemiological data may be incomplete or delayed, limiting the ability to create accurate and up-to-date visualizations [64].
- Data Reliability: The reliability of data is another concern. For example, during disease outbreaks, the rush to report cases can lead to errors in data entry or diagnosis, affecting the accuracy of visualizations. This issue became apparent during the COVID-19 pandemic, where disparities in testing capacities between countries led to underreporting, which distorted global infection maps [65-66].

4.2. Ethical Concerns: Privacy, Data Security, and Misuse

Another major challenge in epidemiological data visualization is managing the ethical concerns surrounding data privacy, security, and the potential misuse of sensitive information. Epidemiological data often includes personal health information, and visualizing this data raises the risk of exposing individuals or groups to privacy violations or discrimination [67].

- Privacy Risks: Geographic visualizations, such as heatmaps or choropleth maps, which display data at local levels, can inadvertently reveal sensitive information about individuals or communities [68]. For example, visualizing HIV/AIDS prevalence in a small town could stigmatize residents, especially if data is not properly anonymized. This breach of privacy can lead to social and legal consequences.
- Data Security: In addition to privacy risks, there are concerns about the security of the data used for visualizations. As epidemiological data increasingly becomes digitized and shared across platforms, the risk of hacking or unauthorized access grows. Data breaches can compromise the confidentiality of patient information and erode public trust in health systems [69-71].
- Misuse of Data: Ethical concerns also extend to how visualized data is interpreted and used. There is the risk that data visualizations could be misused by policymakers or the media to support biased narratives or harmful public health decisions [72]. For example, a misrepresentation of data on vaccine efficacy could be used to fuel anti-vaccine movements, undermining public health initiatives.

To address these ethical concerns, it is crucial for epidemiologists to adhere to strict data governance policies, ensure proper anonymization techniques, and use secure platforms for sharing data visualizations. Transparent communication about the limitations of data can also help prevent misuse.

5. Technological Tools for Data Visualization in Epidemiology

The advancement of technology has significantly enhanced the ability of epidemiologists to visualize and analyze complex datasets. Various tools and platforms have been developed, ranging from statistical software and Geographic Information Systems (GIS) to machine learning (ML) algorithms and open-source platforms for real-time monitoring. This section delves into some of the most widely used technological tools that support data visualization in epidemiology, discussing their capabilities and the role they play in disease surveillance, prediction, and analysis.

5.1. Software and Platforms (e.g., R, Python, ArcGIS, QGIS)

Several specialized software tools are frequently employed in epidemiological research for data visualization. These platforms not only allow researchers to create visual representations of data but also provide powerful analytical functions that help in uncovering patterns and trends in disease spread and health outcomes.

- R: R is a programming language and software environment widely used for statistical computing and graphics. In epidemiology, it is extensively applied for creating visualizations like time series plots, heatmaps, and spatial maps. The rich library of packages, such as ggplot2 and Shiny, enables epidemiologists to develop sophisticated and interactive visualizations [74]. Additionally, R can be integrated with Geographic Information Systems (GIS) for spatial analysis of diseases.
- Python: Like R, Python is another programming language that has gained popularity for data science and visualization. Libraries such as Matplotlib, Seaborn, and Plotly allow users to generate a wide range of visualizations, including histograms, line charts, and 3D plots. Python is highly flexible and integrates well with machine learning libraries, such as scikit-learn and TensorFlow, making it an ideal tool for predictive modeling in epidemiology [75].
- ArcGIS: ArcGIS is a comprehensive Geographic Information System that is widely used for spatial analysis in epidemiology. It enables users to map and analyze disease patterns geographically, using real-world data [76]. Epidemiologists can use ArcGIS to visualize the spread of diseases, assess regional risk factors, and determine the geographical impact of interventions. One notable feature of ArcGIS is its ability to layer different types of data (e.g., population density, environmental factors, healthcare infrastructure) to create complex and informative maps [77].
- QGIS: QGIS is an open-source alternative to ArcGIS and is frequently used for similar spatial visualizations. Although it lacks some of the advanced features of ArcGIS, QGIS is highly customizable, making it a cost-effective option for smaller-scale or resource-limited public health projects [78]. It allows epidemiologists to create choropleth maps, heatmaps, and time-lapse maps that visualize the temporal spread of diseases.

These tools enable epidemiologists to analyze large and diverse datasets and create visualizations that can inform policymakers, health professionals, and the public. Their versatility makes them invaluable in modern epidemiology.

5.2. Machine Learning and AI for Predictive Modeling and Visualizing Complex Datasets

Machine learning (ML) and artificial intelligence (AI) have revolutionized epidemiology by enabling the visualization and analysis of vast, complex datasets that were previously too difficult to interpret [79-81]. Predictive models generated using ML algorithms allow epidemiologists to forecast disease spread, identify high-risk populations, and evaluate the potential impact of public health interventions.

Predictive Modeling: ML algorithms, such as decision trees, random forests, and neural networks, are particularly useful for creating predictive models of disease outbreaks [82]. For instance, these models can take into account numerous variables—such as climate conditions, population mobility, and social behavior—to predict where and when a disease is likely to spread [83]. These predictions can then be visualized using tools like R or Python to show potential future hotspots, enabling preemptive action.

Complex Data Visualization: In addition to predictive modeling, AI can help visualize extremely complex datasets that may involve multiple dimensions, such as genetic data, clinical reports, environmental factors, and social determinants of health [84]. By applying dimensionality reduction techniques, such as principal component analysis (PCA) or t-SNE (t-distributed stochastic neighbor embedding), AI can reduce the complexity of the data and generate visualizations that highlight the most significant variables affecting disease dynamics [85-87].

AI-powered tools are transforming how epidemiologists approach disease monitoring, providing real-time insights that were previously unattainable. However, it is essential to ensure that these models are transparent and interpretable so that they can be trusted by public health authorities and stakeholders.

5.3. Open-Source Tools and Platforms for Real-Time Visualization (e.g., HealthMap, Outbreaks Near Me)

Open-source platforms have become increasingly popular for real-time data visualization and monitoring in epidemiology. These tools often provide access to real-time data from multiple sources and use visualization techniques to present this information in an accessible and informative way [88].

HealthMap: HealthMap is an open-source, real-time platform that aggregates data from various sources, including news reports, social media, and official health alerts, to visualize global disease outbreaks [89]. It uses algorithms to filter and map disease activity, allowing users to see up-to-date information on disease outbreaks worldwide [90]. HealthMap was especially useful during the COVID-19 pandemic, providing real-time information on infection hotspots and trends [91].

Outbreaks Near Me: This tool enables users to track local outbreaks of infectious diseases based on reports from both health authorities and the public. Outbreaks Near Me uses geolocation data to create personalized maps showing nearby disease activity [92]. The platform also incorporates crowd-sourced data, making it a valuable resource for communities seeking real-time updates on local health risks [89].

These platforms demonstrate the power of open-source technology in democratizing access to epidemiological data. They enable both professionals and the public to monitor disease trends, improving community preparedness and response.

6. Conclusion and Recommendations

The use of advanced epidemiological surveillance tools has revolutionised the approach to diseases epidemiology. These technologies enable more accurate predictions, real-time data analysis, and effective decision-making, thereby enhancing the global capacity to detect, respond to, and mitigate outbreaks. With the integration of open-source platforms like R, QGIS, and HealthMap, public health systems are better equipped to handle the complexities of modern health challenges. However, as these technologies evolve, it is necessary to address ethical concerns, data privacy, and the need for enhanced global collaboration to ensure equitable access to these innovations. Based on this, the following recommendations become necessary:

- Strengthening Data Infrastructure: Governments and health organisations should invest in upgrading data collection and storage systems, ensuring they can handle large datasets and support real-time analysis. This will facilitate more accurate disease tracking and prediction.
- Encouraging Cross-Sector Collaboration: Public health bodies, private tech firms, and academic institutions should collaborate to foster innovation in disease surveillance technologies. This partnership will enhance the development of tools that are both effective and widely accessible.
- Ensuring Data Privacy and Ethical Standards: There is a pressing need for stronger regulations and policies to protect sensitive health data. Governments should enact and enforce data privacy laws, ensuring that surveillance systems are used responsibly without compromising individuals' rights.
- Training Public Health Workers: To maximise the potential of AI, machine learning, and big data in epidemiology, public health workers need ongoing training on these new tools. Educational programmes should be established to build capacity in handling advanced surveillance technologies.
- Promoting Public Engagement in Surveillance: Participatory disease surveillance platforms like HealthMap and Outbreaks Near Me have demonstrated the power of crowdsourcing data. Governments should encourage public involvement in these platforms, which can provide valuable real-time data for early outbreak detection.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Li Q, Li Q. Overview of data visualization. Embodying data: Chinese aesthetics, interactive visualization and gaming technologies. 2020:17-47.

[2] Crapo AW, Waisel LB, Wallace WA, Willemain TR. Visualization and the process of modeling: a cognitive-theoretic view. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining 2000 Aug 1 (pp. 218-226).

[3] Munzner T. Visualization. In Fundamentals of computer graphics 2018 Oct 24 (pp. 665-699). AK Peters/CRC Press.

[4] Preim B, Lawonn K. A survey of visual analytics for public health. InComputer Graphics Forum 2020 Feb (Vol. 39, No. 1, pp. 543-580).

[5] Chen H, Zeng D, Yan P, Chen H, Zeng D, Yan P. Data visualization, information dissemination, and alerting. Infectious Disease Informatics: Syndromic Surveillance for Public Health and BioDefense. 2010:73-87.

[6] Carroll LN, Au AP, Detwiler LT, Fu T-c, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: A systematic review. Journal of Biomedical Informatics. 2014;51:287-98.

[7] Park S, Bekemeier B, Flaxman AD. Understanding data use and preference of data visualization for public health professionals: A qualitative study. Public Health Nursing. 2021 Jul;38(4):531-41.

[8] Welhausen CA. Power and authority in disease maps: Visualizing medical cartography through yellow fever mapping. Journal of Business and Technical Communication. 2015 Jul;29(3):257-83.

[9] Tulchinsky TH. John Snow, cholera, the broad street pump; waterborne diseases then and now. Case studies in public health. 2018:77.

[10] BN J. Geographic Information Systems and COVID-19: The Johns Hopkins University Dashboard. Available fom https://europepmc.org/article/ppr/ppr123996

[11] Yang F. Data visualization for health and risk communication. The handbook of applied communication research. 2020 Apr 17:213-32.

[12] Dusse F, Júnior PS, Alves AT, Novais R, Vieira V, Mendonça M. Information visualization for emergency management: A systematic mapping study. Expert Systems with Applications. 2016 Mar 1;45:424-37.

[13] Tsai P, Scott KA, González MC, Pappaioanou M, Keusch GT, editors. Sustaining global surveillance and response to emerging zoonotic diseases..

[14] Zhao IY, Ma YX, Yu MW, Liu J, Dong WN, Pang Q, Lu XQ, Molassiotis A, Holroyd E, Wong CW. Ethics, integrity, and retributions of digital detection surveillance systems for infectious diseases: systematic literature review. Journal of medical Internet research. 2021 Oct 20;23(10):e32328.

[15] World Health Organization. Tools for making good data visualizations: the art of charting. World Health Organization. Regional Office for Europe; 2021. Available from https://apps.who.int/iris/bitstream/handle/10665/342568/WHO-EURO-2021-1998-41753-57181-eng.pdf

[16] Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical Research. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2013. p. 78-79.

[17] Rothman KJ, Greenland S. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 45-48.

[18] Inkelas M, Blair C, Furukawa D, Manuel VG, Malenfant JH, Martin E, Emeruwa I, Kuo T, Arangua L, Robles B, Provost LP. Using control charts to understand community variation in COVID-19. PloS one. 2021 Apr 30;16(4):e0248500.

[19] He Y, Yu X, Gan Y, Zhu T, Xiong S, Peng J, Hu L, Xu G, Yuan X. Bar charts detection and analysis in biomedical literature of PubMed Central. InAMIA Annual Symposium Proceedings 2017 (Vol. 2017, p. 859). American Medical Informatics Association.

[20] Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: time for a new data presentation paradigm. PLoS biology. 2015 Apr 22;13(4):e1002128.

[21] Yadav SK, Akhter Y. Statistical modeling for the prediction of infectious disease dissemination with special reference to COVID-19 spread. Frontiers in public health. 2021 Jun 16;9:645405.

[22] Boscoe FP, McLaughlin C, Schymura MJ, Kielb CL. Visualization of the spatial scan statistic using nested circles. Health & Place. 2003 Sep 1;9(3):273-7.

[23] Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. Journal of biomedical informatics. 2014 Oct 1;51:287-98.

[24] Davenhall WF, Kinabrew C. GIS in health and human services. Springer handbook of geographic information. 2012:557-78.

[25] Hay SI, Omumbo JA, Craig MH, Snow RW. Earth observation, geographic information systems and Plasmodium falciparum malaria in sub-Saharan Africa. Advances in parasitology. 2000 Jan 1;47:173-215.

[26] Netek R, Pour T, Slezakova R. Implementation of heat maps in geographical information system–exploratory study on traffic accident data. Open Geosciences. 2018 Aug 15;10(1):367-84.

[27] Maciejewski R, Rudolph S, Hafen R, Abusalah A, Yakout M, Ouzzani M, Cleveland WS, Grannis SJ, Ebert DS. A visual analytics approach to understanding spatiotemporal hotspots. IEEE Transactions on Visualization and Computer Graphics. 2009 Aug 28;16(2):205-20.

[28] Huang Z. Spatiotemporal evolution patterns of the COVID-19 pandemic using space-time aggregation and spatial statistics: a global perspective. ISPRS International Journal of Geo-Information. 2021 Aug;10(8):519.

[29] Liu Y, Gu Z, Liu J. Uncovering transmission patterns of COVID-19 outbreaks: A region-wide comprehensive retrospective study in Hong Kong. EClinicalMedicine. 2021 Jun 1;36.

[30] Osei F. Current statistical methods for spatial epidemiology: a review. Austin Biometrics Biostat. 2014 Jan;1(2):1-7.

[31] Selvaratnam I. Disease Maps and Geographic Map Projections-A Simulation Study and Scoping Review. Available from https://atrium.lib.uoguelph.ca/bitstreams/60575cb5-641b-4ba5-b0b4-747b5f1d85f3/download

[32] Waruru A, Achia TN, Tobias JL, Mwangi M, Wamicwe J, Zielinski-Gutierrez E, Oluoch T, Muthama E, Tylleskär T. Finding hidden HIV clusters to support geographic-oriented HIV interventions in Kenya. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2018 Jun 1;78(2):144-54.

[33] Park J, Choi J, Choi JY. Network analysis in systems epidemiology. Journal of Preventive Medicine and Public Health. 2021 Jul;54(4):259.

[34] Luke DA, Harris JK. Network analysis in public health: history, methods, and applications. Annual review of public health. 2007 Apr 21;28(1):69-93.

[35] Riley S. Large-scale spatial-transmission models of infectious disease. Science. 2007 Jun 1;316(5829):1298-301.

[36] Zhang Y, Sun Y, Padilla L, Barua S, Bertini E, Parker AG. Mapping the landscape of COVID-19 crisis visualizations. InProceedings of the 2021 CHI conference on human factors in computing systems 2021 May 6 (pp. 1-23).

[37] Aldila D, Khoshnaw SHA, Safitri E, Anwar YR, Bakry ARQ, Samiadji BM, et al. A mathematical study on the spread of COVID-19 considering social distancing and rapid assessment: The case of Jakarta, Indonesia. Chaos, Solitons & Fractals. 2020;139:110042.

[38] Kumar N, Susan S. COVID-19 pandemic prediction using time series forecasting models. In2020 11th international conference on computing, communication and networking technologies (ICCCNT) 2020 Jul 1 (pp. 1-7). IEEE.

[39] Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. International journal of epidemiology. 2013 Aug 1;42(4):1187-95.

[40] Erbas B, Hyndman RJ. Data visualisation for time series in environmental epidemiology. Journal of Epidemiology and Biostatistics. 2001 Nov 1;6(6):433-43.

[41] Rakocevic B, Grgurevic A, Trajkovic G, Mugosa B, Grujicic SS, Medenica S, Bojovic O, Alonso JE, Vega T. Influenza surveillance: determining the epidemic threshold for influenza by using the Moving Epidemic Method (MEM), Montenegro, 2010/11 to 2017/18 influenza seasons. Eurosurveillance. 2019 Mar 21;24(12):1800042.

[42] Andrienko N, Andrienko G. Cumulative curves for exploration of demographic data: a case study of Northwest England. Computational Statistics. 2004 Feb;19(1):9-28.

[43] Washburne A, Silverman J, Lourenco J, Hupert N. Analysis and visualization of epidemics on the timescale of burden: derivation and application of Epidemic Resistance Lines (ERLs) to COVID-19 outbreaks in the US. medRxiv. 2021 May 5:2021-05.

[44] Martinez ME. The calendar of epidemics: Seasonal cycles of infectious diseases. PLoS pathogens. 2018 Nov 8;14(11):e1007327.

[45] Wieser A, Reuss F, Niamir A, Müller R, O'Hara RB, Pfenninger M. Modelling seasonal dynamics, population stability, and pest control in Aedes japonicus japonicus (Diptera: Culicidae). Parasites & vectors. 2019 Dec;12:1-2.

[46] Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. JMIR human factors. 2018 May 31;5(2):e9328..

[47] Batt S, Grealis T, Harmon O, Tomolonis P. Learning Tableau: A data visualization tool. The Journal of Economic Education. 2020 Aug 17;51(3-4):317-28.

[48] Akhtar N, Tabassum N, Perwej A, Perwej Y. Data analytics and visualization using Tableau utilitarian for COVID-19 (Coronavirus). Global Journal of Engineering and Technology Advances. 2020 May 25.

[49] Dixon BE, Grannis SJ, McAndrews C, Broyles AA, Mikels-Carrasco W, Wiensch A, Williams JL, Tachinardi U, Embi PJ. Leveraging data visualization and a statewide health information exchange to support COVID-19 surveillance and response: application of public health informatics. Journal of the American Medical Informatics Association. 2021 Jul 1;28(7):1363-73.

[50] Graves SM, He L. Covid-19 mapping with microsoft power bi. Terra Digitalis. 2020 Oct 31:1-5.

[51] Mathers CD. History of global burden of disease assessment at the World Health Organization. Archives of Public Health. 2020 Dec;78:1-3.

[52] Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. International journal of health geographics. 2020 Dec;19(1):1-2.

[53] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. jama. 2020 Apr 7;323(13):1239-42.

[54] Kho AN, Pacheco JA, Peissig PL, Rasmussen LV, Newton KM, Weston N, et al. Electronic health records for genetic research: Results from the eMERGE network. Science translational medicine. 2011;3(79):79re1.

[55] McGinnis DF, Ferguson RP. Electronic health records and health care quality. American journal of medical quality. 2011;26(4):272-9.

[56] Martínez-Gamboa RA, Lozano-Rodríguez JC, Hernández-García C, Del Moral-Damián MP, Andrade-Rosa EI, Morales-García MS. Laboratory information management system to enhance tracking and analysis of clinical laboratory results. Journal of laboratory automation. 2015;20(6):662-9.

[57] Boolchandani M, D'Souza AW, Dantas G. Sequencing-based methods and resources to study antimicrobial resistance. Nature Reviews Genetics. 2019 Jun;20(6):356-70.

[58] O'Sullivan D, Rahamathulla M, Pawar M. The impact and implications of COVID-19: An Australian perspective. The International Journal of Community and Social Development. 2020;2(2):134-51.

[59] Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, et al. Aggregated mobility data could help fight COVID-19. Science. 2020;368(6487):145-6.

[60] Kumar A, Sharma K, Singh H. Real-time epidemic monitoring using social media data. Informatics in medicine unlocked. 2020;20:100403.

[61] Kovats RS, Campbell-Lendrum DH, McMichael AJ, Woodward A, Cox J. Early effects of climate change: Do they include changes in vector-borne disease? Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 2001;356(1411):1057-68.

[62] Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al. A global map of dominant malaria vectors. Parasites & vectors. 2012;5(1):1-11.

[63] Epstein PR. Climate change and human health. New England journal of medicine. 2005;353(14):1433-6.

[64] Wetherill O, Lee CW, Dietz V. Root causes of poor immunisation data quality and proven interventions: a systematic literature review. Annals of infectious disease and epidemiology. 2017 Feb 2;2(1):1.

[65] Huber SG, Helm C. COVID-19 and schooling: evaluation, assessment and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. Educational Assessment, Evaluation and Accountability. 2020;32(2):237-70.

[66] Nkengasong J, Iwasaki A, Victora C, Oh J, Gao GF, Agrawal A, Drosten C, Söderberg-Naucler C, López-Collazo E, Pollock AM, Viola A. The global response to the COVID-19 pandemic. Med. 2020 Dec 18;1(1):3-8.

[67] Laney D. 3D data management: Controlling data volume, velocity, and variety. META Group Research Note. 2020;6(70):1-4.

[68] Cromley RG, Cromley EK. Choropleth map legend design for visualizing community health disparities. International journal of health geographics. 2009 Dec;8:1-1.

[69] Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. Journal of the American Medical Informatics Association. 2011 Jul 1;18(4):515-22.

[70] Fernández-Alemán JL, Señor IC, Lozoya PÁO, Toval A. Security and privacy in electronic health records: A systematic literature review. Journal of Biomedical Informatics. 2013;46(3):541-62.

[71] Seh AH, Zarour M, Alenezi M, Sarkar AK, Agrawal A, Kumar R, Ahmad Khan R. Healthcare data breaches: insights and implications. In Healthcare 2020 May 13 (Vol. 8, No. 2, p. 133). MDPI.

[72] Howe III EG, Elenberg F. Ethical challenges posed by big data. Innovations in clinical neuroscience. 2020 Oct;17(10-12):24.

[73] Correll M. Ethical dimensions of visualization research. In Proceedings of the 2019 CHI conference on human factors in computing systems 2019 May 2 (pp. 1-13).

[74] Ponce M, Sandhel A. covid19. analytics: An R package to obtain, analyze and visualize data from the Coronavirus disease pandemic. arXiv preprint arXiv:2009.01091. 2020 Sep 2.

[75] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011 Nov 1;12:2825-30.

[76] Flies EJ, Williams CR, Weinstein P, Anderson SJ. Improving public health intervention for mosquito-borne disease: the value of geovisualization using source of infection and LandScan data. Epidemiology & Infection. 2016 Oct;144(14):3108-19.

[77] Khashoggi BF, Murad A. Issues of healthcare planning and GIS: a review. ISPRS International Journal of Geo-Information. 2020 May 27;9(6):352.

[78] Black B. Remote Sensing-Based Coastal Environmental Monitoring Approaches for Remote Island Settings. Available from https://ir.library.oregonstate.edu/downloads/mp48sn97h

[79] Panesar A. Machine learning and AI for healthcare. Coventry, UK: Apress; 2019.

[80] Williams AM, Liu Y, Regner KR, Jotterand F, Liu P, Liang M. Artificial intelligence, physiological genomics, and precision medicine. Physiological Genomics. 2018;50(4):237-43.

[81] Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. The American journal of medicine. 2019 Jul 1;132(7):795-801.

[82] Alfred R, Obit JH. The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review. Heliyon. 2021 Jun 1;7(6).

[83] Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. Covid-19 outbreak prediction with machine learning. Algorithms. 2020 Oct 1;13(10):249.

[84] Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. Genome medicine. 2019 Nov 19;11(1):70.

[85] Birjandtalab J, James M, Nourani M, Harvey J. Learning from non-seizure clusters for eeg analytics. In2018 IEEE Biomedical Circuits and Systems Conference (BioCAS) 2018 Oct 17 (pp. 1-4). IEEE.

[86] Liebal UW, Phan ANT, Sudhakar M, Raman K, Blank LM. Machine Learning Applications for Mass Spectrometry-Based Metabolomics. Metabolites. 2020; 10(6):243. https://doi.org/10.3390/metabo10060243.

[87] Khalid S, Khalil T, Nasreen S. A survey of feature selection and feature extraction techniques in machine learning. In2014 science and information conference 2014 Aug 27 (pp. 372-378). IEEE.

[88] Chorianopoulos K, Talvis K. Flutrack. org: Open-source and linked data for epidemiology. Health informatics journal. 2016 Dec;22(4):962-74.

[89] Freifeld C. *Participatory epidemiology: harnessing the HealthMap platform for community-based disease outbreak monitoring* (Doctoral dissertation, Massachusetts Institute of Technology). Available from https://dspace.mit.edu/bitstream/handle/1721.1/57707/656281875-MIT.pdf?sequence=2

[90] Bruno P, Calimeri F. Using heatmaps for deep learning based disease classification. In2019 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB) 2019 Jul 9 (pp. 1-7). IEEE.

[91] Caspi G, Chen J, Liverant-Taub S, Shina A, Caspi O. Heat maps for surveillance and prevention of COVID-19 spread in nursing homes and assisted living facilities. Journal of the American Medical Directors Association. 2020 Jul;21(7):986.

[92] Arias-Carrasco R, Giddaluru J, Cardozo LE, Martins F, Maracaja-Coutinho V, Nakaya HI. OUTBREAK: a user-friendly georeferencing online tool for disease surveillance. Biological Research. 2021 Dec;54:1-6.