International Journal of Advanced Research in Engineering Technology & Science ISSN: 2349-2819
Impact Factor: 7.10
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Email: editor@ijarets.org Volume-11 Issue-12 December-2024 www.ijarets.org

"PUBLIC HEALTH SURVEILLANCE SYSTEMS: USING BIG DATA ANALYTICS TO PREDICT INFECTIOUS DISEASE OUTBREAKS"

Sheraz Fatima
Indiana Wesleyan University, Marion Indiana
Email ID: sherazfatima@gmail.com

ABSTRACT

The increasing trends in the number of global health threats prompt the importance of effective Public Health Surveillance. Thus, this work aims at identifying how big data analytics can be used to predict Epidemic infectious diseases based on historical and real-time data from the social media, EHRs, as well as environmental surveillance systems. Diagnostic models were built for pattern recognition and anomaly detection to predict likely areas for diseases such as COVID-19, dengue, and influenza. This research shows that the implementation of big data analytics into public health care systems raises preventive approach and offers efficient control measures for managing outbreaks, which weakens these influences' consequences. This paper also gives the difficulties of data integration, privacy issues and scalability for big data surveillance system in the recommended plan for big data surveillance across the world.

Keywords: Big Data Analytics, Infectious Disease Outbreaks, Public Health Surveillance, Machine Learning, Early Detection

Introduction

The appearance of different types of diseases can lead to massive infections and deaths, which remain a threat to the world's health care systems. The irregular and dynamic feature of such outbreaks make it crucial to identify such illnesses early and follow up with the necessary intervention. A common type of surveillance that has been widely implemented in the healthcare industry functions well in many ways but falls short in delivering timely or proactive information about an emerging epidemic. The advent of big data as large data sets originating from numerous sources inclusive of social media, Electronic Health Records (EHR) and environmental monitoring systems is an opportunity to improve on public health surveillance.

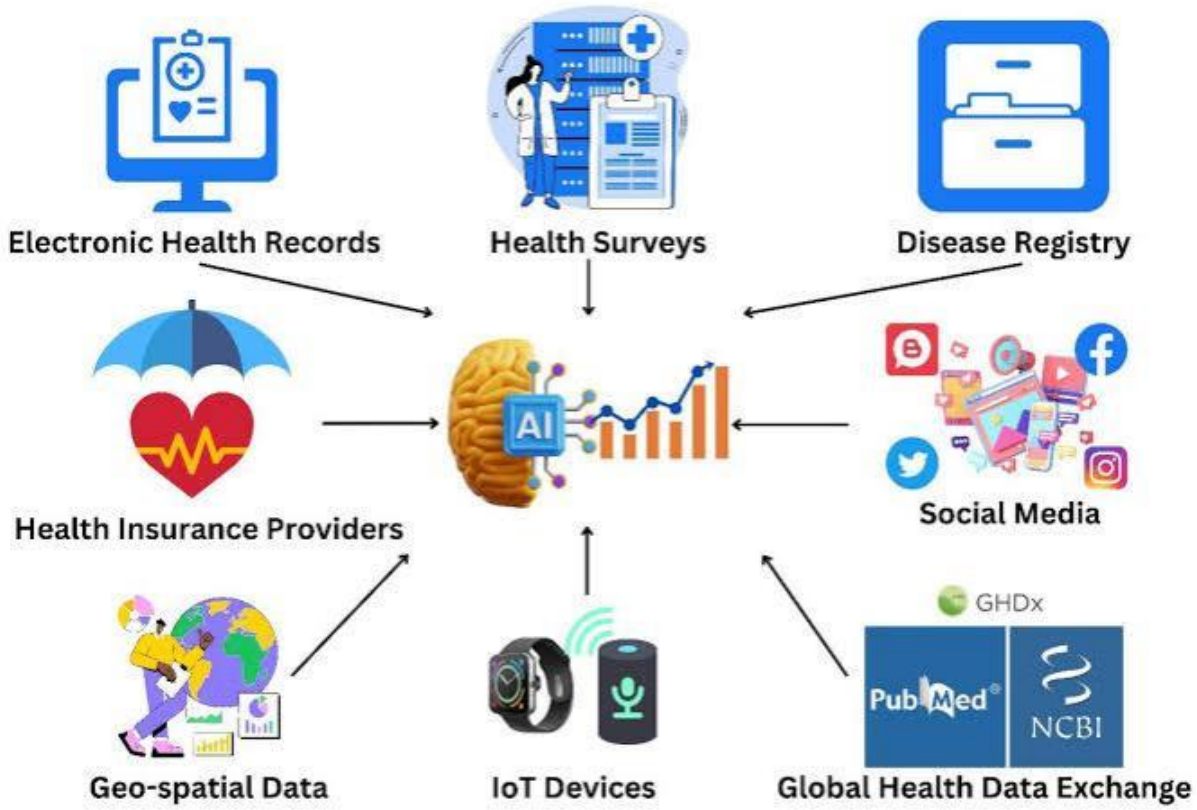

Figure 1: Artificial Intelligence in Public Health: Revolutionizing (Pranav et al., 2023).

The advent of big data also holds the promise of radically changing the way we might identify, anticipate, and address contagious emerging diseases through machine learning (ML) and artificial intelligence (AI). In this approach, using massive data from different sources, organisations within the public health domain can have some insights into such outbreaks at their early stages, as well as having an opportunity to determine areas where outbreaks are most probable and provide necessary resources in the shortest time possible. The study of this paper is to analyse how big data analytics can be used in predicting an outbreak of infectious diseases such as COVID 19, dengue and influenza. An additional object of discussion will also include the issues and hitches that come with the implementation of these technologies into the prevailing structures of public health domains.

Background

Public health surveillance may be defined as the continuous, systematic and systematic identification, collection, analysis and interpretation of health-related data for public health purposes. Historically, physicians' reports, laboratory testing, and record review provide the basis of safety surveillance. The problem with most of these systems is that they can take a while before they can capture the data and the report back on an epidemic (BotPenguin, 2024).

Scholars start to focus on big data analytics as one of the promising developments to improve the performance of surveillance systems. Big data on the other hand is a complex, massive volume of data and

high velocity data that cannot be managed and analysed by traditional techniques. Such datasets may contain numbers, texts, bio and social media data, and even the coordinates of geolocation from mobile devices (Chae et al., 2018). The use of such heterogenous data streams presents a novel chance to identify emerging trends and forecast epidemics in progress.

For instance, in the 2014-2016 Ebola outbreak in West Africa, the researcher used data derived from mobile phone to describe mobility patterns in surveillance of high-risk populations for the diseases (Pranav Anjaria et al., 2023). The same is the case with other social media platforms where Twitter has also been used to track such disease debates to identify emerging signs of the diseases. Through main estimation of these data, it is possible to predict the probability of an outbreak in certain regions based on certain parameters using machine learning algorithms.

Prediction of Infectious Disease using Big Data Analytics

One of the techniques under big data analytics includes data mining, machine learning as well as statistical modelling that can be used to identify the probability of spread of infectious diseases (Jia et al., 2020). Through comparing current trends and outage of specific diseases, the official of public health can detect risks and threats to prevent them.

- Social Media Monitoring

Various social networks act as effective sources of analysing trends related to public health. Information that concerns health issues and diseases, symptoms, existing concerns of the public and general behavior can be found in posts on social media including Twitter, face book and even Reddit. It noted that data mining of social media data can help researchers to detect warning signs of outbreak (Igwama et al., 2024). For example, when discussing flu-like symptoms on social media increases it is most likely that flu has started spreading in its early stages.

- For Electronic Health Records (EHRs)

EHRs contain extensive information of patients, including symptoms, diagnosis and treatment outcomes. Using machine learning models on EHR data, PHOs can observe hitherto unseen patterns of morbidity and infectious disease outbreaks. For example, increase in daily hospitalization rate of respiratory illnesses may signal flu season therefore authorities can act accordingly (Chowdhury et al., 2024).

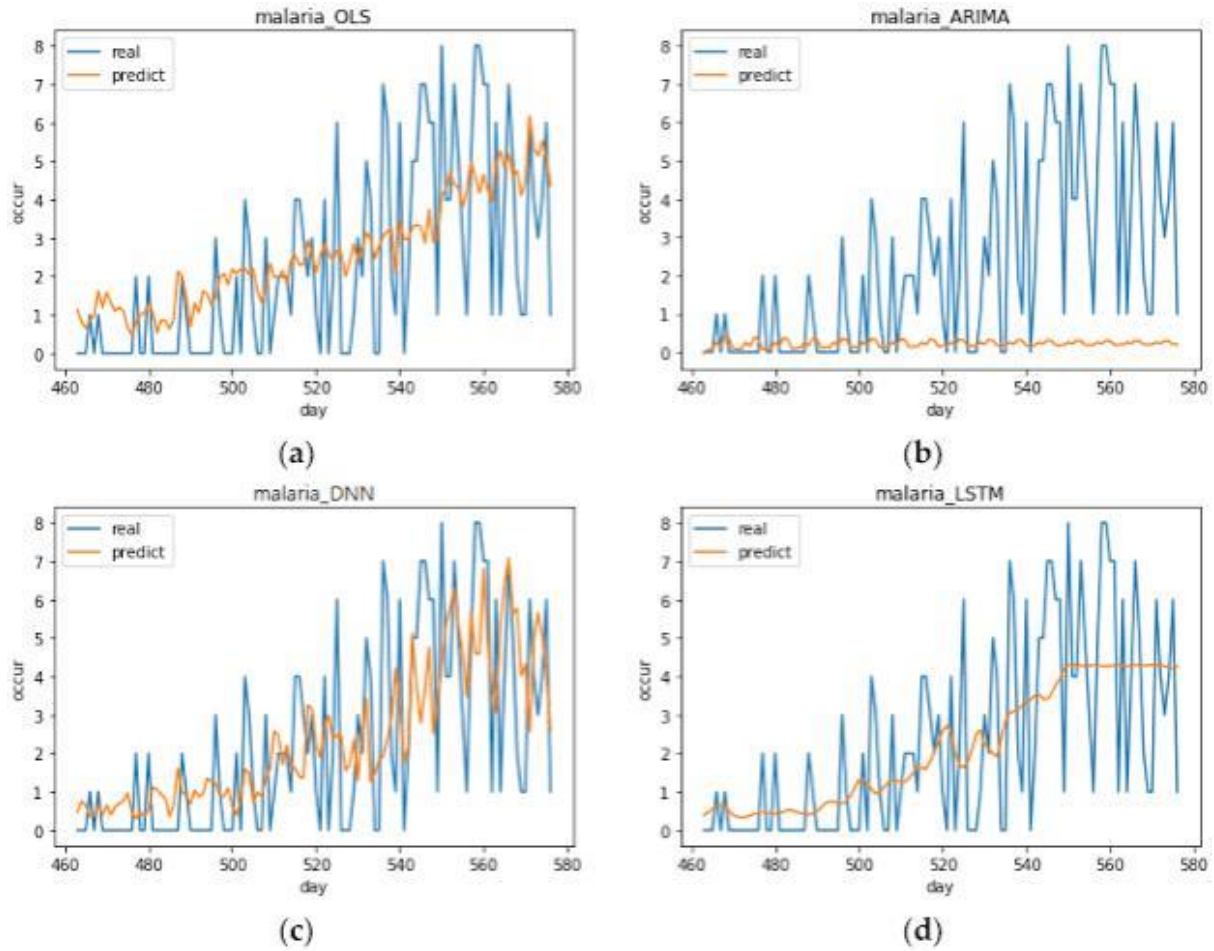

Figure 2: Predicting Infectious Disease Using Deep Learning and Big Data (Chae et al., 2018).

- Environmental Monitoring Systems

Other health related data could be including climatic factors like; temperature, humidity, and air quality control can also help to determine diseases distribution. Several communicable diseases for instance ; dengue and malaria are known to be sensitive to the physical environment (Khodadadi & Towfek, 2023). When health data is combined with environmental data, epidemiologists and other researchers can model future ebbs and flows of disease occurrence based on the changing environment.

- Mobile Health (mHealth) Data

Smartphones and now wearable technologies have taken central stage in monitoring health related data. These devices can capture information on these indicators and sign-containing symptoms, physical activity levels, as well as potential environmental exposures. When such data is collected from a large population, it helps a researcher make early signs that denote the onset of an infectious disease.

Big Data Integration Analysis of the Challenges

While big data analytics offers significant promise in enhancing public health surveillance, several challenges must be addressed to fully realize its potential:

- Data harmonisation and combination

Several of the authors' concerns for using big data for public health surveillance revolve around the issue of data fusion (Ahmed et al., 2021). Usually, the health data remain in different formats and in different systems as well as the environmental data or the data obtained from social networks. It is crucial to get a uniform structure and make sure that data are compatible for various organizational systems when developing elaborate forecasting models.

Figure 3: Challenges of Big Data (Botpenguin, 2024).

- Privacy and Security Concerns

Personal health data as well as utilization of social media information create major issues of privacy and security. Therefore, public health organizations must meet legal requirements concerning data protection, for example, the GDPR in Europe, while making that data usable for analytics. One of the main issues that must be addressed is how privacy can be reconciled with the requirement of generating data on time (Zeng et al., 2021).

- Scalability and infrastructure

Big data analytics is computation intensive and characterized by a need for very large computational platforms to handle big data. There are also many public health organizations that especially those from low and middle-income countries do not possess adequate infrastructural support to conduct the big data surveillance systems. Those barriers call for investments in scalable technologies and cloud-based platforms to spur stable technology advances.

- Data Quality and Accuracy

The reliability of predictions made from big data models can therefore be largely attributed to the kind of data being used (Munagandla et al., 2024). The problem of getting wrong data is that it often results in a wrong forecast and erroneous analyses. Accuracy when collecting data and reducing errors is essential in the success of big data analytics in public health surveillance.

Case Studies

Several case studies illustrate the successful application of big data analytics in predicting infectious disease outbreaks:

- COVID-19 Prediction Via social media and Electronic Health Records

In the COVID-19 situation, EHR and social media data were adopted for the analysis of potential upcoming and future progress of COVID-19. Early warning systems grew from the analysis of messages containing COVID symptoms by monitoring the Twitter data and incorporating the EHR data to map the prevalence of the virus (BotPenguin, 2024). This early detection allowed the authorities to conduct targeted quarantine measures including local lock down or prevented people from travelling to certain areas.

- Dengue Fever Forecasting

In areas where dengue fever is prevalent big data analytics is some of the few methods which have been employed to estimate the disease. The climate data include Rainfall patterns, temperature combined with past case records from hospital were used to generate models that largely accurately predicted dengue fever outbreaks (Chae et al., 2018). These models helped the public health authorities to intervene in early stages using things like control of the mosquitoes before the disease got to these extreme levels.

1. Influenza Surveillance using EHRs and Mobile Data

In the United States, through the Centres for Disease Control and Prevention, popular instance of EHR and mHealth use is to track the flu (Pranav Anjaria et al., 2023). This way the CDC will be able to track the flu symptoms from the data of hospitals and Health apps in real time and thereby estimate probable flu affected areas to allocate resources and respond more promptly.

Conclusion

The sample shows that the incorporation of big data analytical tools into public health monitoring networks is a revolution in combating infectious diseases. Real-time information from various sources enables public health organizations enhance forecasting and management of such events hence reducing their impact on lives and the economy. At the same time, issues of integration, privacy, scalability, and data quality remain critical to obtain the maximum beneficial effect of these systems. Hence, protecting world health public as new technologies enhance the development of the future work of public health surveillance this will prominently rely on big data.

References

BotPenguin. (2024, June 12). *Challenges of Big Data*. Botpenguin.com; BotPenguin. https://botpenguin.com/glossary/challenges-of-big-data

Chae, S., Kwon, S., & Lee, D. (2018). Predicting Infectious Disease Using Deep Learning and Big Data. *International Journal of Environmental Research and Public Health*, *15*(8), 1596. https://doi.org/10.3390/ijerph15081596

Pranav Anjaria, Varun Asediya, Bhavsar, P., Pathak, A., Desai, D. N., & Patil, V. (2023). Artificial Intelligence in Public Health: Revolutionizing Epidemiological Surveillance for Pandemic Preparedness and Equitable Vaccine Access. *Vaccines*, *11*(7), 1154–1154. https://doi.org/10.3390/vaccines11071154

Jia, Q., Guo, Y., Wang, G., & Barnes, S. J. (2020). Big data analytics in the fight against major public health incidents (Including COVID-19): a conceptual framework. *International journal of environmental research and public health*, *17*(17), 6161.

Igwama, G. T., Olaboye, J. A., Maha, C. C., Ajegbile, M. D., & Abdul, S. (2024). Big data analytics for epidemic forecasting: Policy Frameworks and technical approaches. *International Journal of Applied Research in Social Sciences*, *6*(7), 1449-1460.

Chowdhury, A. T., Newaz, M., Saha, P., Majid, M. E., Mushtak, A., & Kabir, M. A. (2024). Application of Big Data in Infectious Disease Surveillance: Contemporary Challenges and Solutions. In *Surveillance, Prevention, and Control of Infectious Diseases: An AI Perspective* (pp. 51-71). Cham: Springer Nature Switzerland.

Khodadadi, E., & Towfek, S. K. (2023). Internet of Things Enabled Disease Outbreak Detection: A Predictive Modeling System. *Journal of Intelligent Systems & Internet of Things*, *10*(1).

Ahmed, I., Ahmad, M., Jeon, G., & Piccialli, F. (2021). A framework for pandemic prediction using big data analytics. *Big Data Research*, *25*, 100190.

Zeng, D., Cao, Z., & Neill, D. B. (2021). Artificial intelligence–enabled public health surveillance—from local detection to global epidemic monitoring and control. In *Artificial intelligence in medicine* (pp. 437-453). Academic Press.

Munagandla, V. B., Dandyala, S. S. V., & Vadde, B. C. (2024). AI-powered cloud-based epidemic surveillance system: A framework for early detection. *Revista de Inteligencia Artificial en Medicina*, *15*(1), 673-690.