EURASIAN EXPERIMENT JOURNAL OF PUBLIC HEALTH (EEJPH)

ISSN: 2992-4081 ©EEJPH Publications

Volume 7 Issue 2 2025

Artificial Intelligence in Predicting Epidemic Outbreaks

Kungu Erisa

Department of Pharmacognosy Kampala International University Uganda
Email: erisa.kungu@studwc.kiu.ac.ug

ABSTRACT

Epidemic outbreaks have historically posed significant public health challenges, causing substantial morbidity, mortality, and socio-economic impacts. Early prediction is important for implementing timely interventions to mitigate these effects. Artificial intelligence (AI) has emerged as a transformative tool in predicting epidemic outbreaks, leveraging machine learning, deep learning, and bioinformatics to analyze vast datasets and identify patterns. This paper examines the role of AI in forecasting epidemics, focusing on the potential of various algorithms, data sources, and real-world applications. It also addresses challenges such as data quality, model bias, interdisciplinary collaboration, and ethical concerns. By integrating traditional epidemiological approaches with AI-driven methods, we propose strategies for enhancing prediction accuracy and improving public health preparedness. The study underscores the need for robust regulatory frameworks and equitable deployment of AI technologies in epidemic prediction.

Keywords: Artificial Intelligence, Epidemic Prediction, Machine Learning, Public Health, Data Analytics, Ethical Considerations.

INTRODUCTION

Epidemic outbreaks are an important public health problem, and the periodic emergence of infectious epidemics continues to cause an increased toll of morbidity and mortality in affected populations. The appearance of modern transportation routes on a global scale has facilitated the spread of pathogens across international borders. An important aspect of controlling and mitigating the spread of infectious epidemics includes the possibility of predicting them as early as possible. Only in this way can adequate monitoring and control strategies be implemented to prevent long-term propagation among the population, along with the socio-economic damage that always accompanies unpredictable and uncontrolled epidemics. Historical records show that successful prediction and timely intervention have had a substantial impact on the spread of an infectious outbreak. Unpredicted diseases, especially those due to the reactivation or emergence of new viral mutants, have historically had devastating effects on human populations. The recent outbreak of Ebola virus disease in West Africa led to numerous cases and recorded deaths. One of the main reasons for this tragedy was the lack of adequate and organized early intervention measures. Another more recent example is the COVID-19 pandemic, resulting in a significant number of confirmed cases and deaths worldwide. Despite the presence of a health care system, the progress of the infectious outbreak could not be adequately controlled without a realignment of the public health strategies underlying such an epidemic, such that the system was overwhelmed by the population of infected and symptomatic patients. The development of digital surveillance methods and the collection of new clinical data to feed prediction models based on artificial intelligence could be of great help in the future [1, 2].

Overview of Epidemics and Their Impact

Epidemics occur when disease cases in a community exceed typical expectations. A distinction is usually made at the country level; higher incidence or emerging diseases necessitate confirmation as epidemics. For instance, myxomatosis in Germany in 1999 was considered an epidemic due to inadequate responses in managing cases. Diseases can evolve into severe epidemics threatening countries or global levels, significantly impacting local communities and economies. Throughout history, epidemics have stalled or reversed human development, and modern pandemics pose new political challenges, including border closures. Predictors suggest modest epidemic control linked to economic growth, while historical analyses reveal that political and economic structures influence a nation's preparedness to respond. Between 2000 and 2017, 14 epidemics in the Pacific region, resulting from a buildup of susceptible individuals, were treated as significant due to their economic impacts. Responses should parallel those to Avian Influenza, which strained the global economy despite perceptions of seasonal flu's higher infection rates. Current interventions are limited by small sample sizes due to rapidly changing epidemic conditions, yet past epidemic data can enhance intervention strategies. Effective epidemic control necessitates addressing overall transmission dynamics rather than solely focusing on treatment resources. Predicting epidemic patterns is crucial for timely and adequate responses, encompassing predicting initiation, size, spread, severity, and recovery rates. Accurate predictions can enable tailored public health measures and economic adjustments. For instance, understanding epidemic spread can inform workforce production strategies, as evidenced by hierarchical regression analysis during previous flu outbreaks [3, 4].

Role of Artificial Intelligence in Epidemic Prediction

Artificial intelligence (AI) is revolutionizing the prediction of epidemic situations. The ability to process large datasets is the greatest impression AI has left in every study, especially for epidemiologists who used to journey through multiple datasets during outbreak situations. Data analysis tools like machine learning, deep learning, predictive analytics, and bioinformatics are bringing cutting-edge outcomes by identifying hidden patterns. One of the best advantages of using AI for epidemic prediction is that these algorithms work on earlier outbreak datasets, learn from them, and predict future outbreak results with more precision. Currently, using AI models, scientists are outperforming classical mathematical models in forecasting outcomes during outbreaks. It has been witnessed from real-time operation centers that AI could early predict outbreaks, preventing wider spread [5, 6]. While AI in disease outbreaks has brought new light to the field of public health, the discipline is not an alternative to traditional methods; instead, it is a mere innovation. For a hundred years, long before AI was born, the interdisciplinary approach with the participation of clinicians, epidemiologists, laboratory diagnostic levels, statisticians, virologists, immunologists, and biologists derived the concept of One Health/EcoHealth/Clinical Informatics. AI-based disease outbreak prediction is the amalgamation of computer science and epidemiological research, whose goal is to improve the ability to predict, detect, and control infectious diseases more efficiently. Its main intent is to predict outbreaks before secondary cases and mortalities occur, which might happen in the next three months within a limited area. As the pandemic began, scientists showed increased interest in understanding the algorithmic and mathematical aspects of AI and their impact on infectious diseases. Despite rapid advancements in AI, adapting this specialized tool to the needs of epidemic predictions has not been straightforward [7, 8].

Machine Learning Algorithms for Epidemic Prediction

We can find several machine learning algorithms that can be used in modeling and predicting epidemic outbreaks. From the application point of view, we can divide these algorithms into three groups: 1. Classification algorithm: These algorithms can be used to divide units based on class labels into groups. They can categorize units into categories such as infected and susceptible, risk group and normal group, etc. 2. Regression algorithm: These algorithms can be used to predict exact numbers and quantities. They are used to forecast new infection cases after several days or months, mortality prediction, infectious reservoir prediction, etc. 3. Clustering-based algorithm: Clustering is an unsupervised learning method. It can be used to classify community units into several groups. This method is frequently used to provide insights into underlying clusters of units that are sometimes not clear in the data. An approach that can use two or three algorithms is very important because prediction, knowing the future situation, is a fundamental step to developing simulation and decision support. The prediction uses past data and future data. The supervised method uses past data to predict future data over time. Choosing the appropriate learning algorithm that has the proper natural capability to represent the structure of the time series is the key to accurate time-series prediction. Real disease cases show that the prediction performance of machine learning is quite good, but the coverage is still below that of the deterministic model, which has a high-intensity computation cost. The limitation of the prediction performance of machine learning algorithms is due to a poor selection of the most relevant features, a lack of understanding of machine learning algorithms, and imperfect tuning and validation in the model [9, 10].

Data Sources and Collection Methods

Data play critical roles in predicting outbreaks. The disease transmission process and the ground situation of epidemics can be reflected by the data. There are different sources of data. Traditional data sources are mainly shaped by the healthcare system. Hospitals and health departments are the primary sources of information. Non-traditional data come from various sources. During an emergence, incapacitated people, shelter-seeking individuals, and willing resources appear. Information about the emergent disaster can be reflected by social media and mobile technology [11, 12]. We discuss several pros and cons of the data sources. Timeliness and representativeness are essential for data, especially commercial data. However, data are hard to collect due to several reasons, including computational difficulty and privacy concerns. This is a critical connection between data and epidemic predictive modeling. The investigation relates to the methodology used for data collection. Interestingly, there are many suitable data collection methods. At the same time, there are methods to preprocess the raw collected data as well. The features of the data should be extracted. The data should be washed, formatted, and cleaned. More importantly, the data should be integrated and unified in line with different sources from various methods. It is necessary to maintain the flow of continuous data. Daily updated and continuous data streams are essential for forecasting their progression. However, more work has to be done to ensure data flow in real-time. Data integration is necessary for model making. The data should reflect the ground situation, which coincides with the essential characteristics of the epidemic model under study [13, 14].

Traditional Data Sources Vs. Non-Traditional Data Sources

Despite initial success in outbreak prediction, anticipating patterns and new threats remains challenging. Early detection failures may stem from traditional infectious disease surveillance, which has limited predictive value. Recently, non-traditional data sources, like news and internet searches, gained attention for early outbreak detection. While traditional methods rely on structured healthcare data, non-traditional sources gather extensive, real-time data across various demographics, enabling public engagement in health issues and monitoring health communication efforts. One advantage of non-traditional data is the rapid tracking of new outbreaks, especially since the internet reflects public concerns about communicable diseases. Algorithms now monitor keywords in discussions and searches, identifying spikes that may indicate early outbreaks. However, these data often reference various events and include public health emergencies, complicating connections to non-English-speaking populations. The media's role in raising awareness of outbreaks can be seen as 'predictable.' However, timely data successfully indicated a dengue fever outbreak in five Latin American countries two weeks in advance. A new methodology has been created to utilize internet data for this kind of tracking, incorporating natural language processing, as traditional monitoring neglects how people search for information. Despite this, data accuracy remains a concern, as they are often incomplete compared to official statistics. Thus, finding a balance of coverage, accuracy, and timeliness for optimal early warning is crucial. In some instances, integrating traditional and non-traditional strategies can provide better predictions; for instance, a multidisciplinary social-biological approach uses biological, ecological, and socio-economic data alongside slow-official notifications, such as influenza case reports [15, 16].

Case Studies and Applications

Recent studies highlight the increasing use of machine learning and artificial intelligence in epidemic prediction. The flu's rapid spread has led to a new method utilizing recurrent neural networks for forecasting. A novel neural network approach using historical influenza-like illness trends at the state level showed significant precision. Researchers, using several LSTM-based methods, predicted infections from influenza A and B, as well as ILI activity in cities across South China and Indonesia. Additionally, LSTM-based models effectively forecasted pediatric dengue fever. A convolutional neural network was also employed to predict new COVID-19 cases. While some projects remain unpublished, a team is currently evaluating event-based alarm scores. Moreover, refining data usage and coverage selection for biosurveillance algorithms poses challenges in public health deployment [17, 18].

Real-World Examples of AI in Epidemic Prediction

Which real-world applications are actually available right now? Here, we introduce several projects and associated technologies designed to predict outbreaks, summarizing them in a single list, and exploring successes, failures, and other considerations for each. We pay close attention to new 'case study' applications launched at the onset of the pandemic. While considering the case studies themselves, we also examine more abstractly what it means to make decisions from data, and the broader impacts and policy relevance of the approaches that failed and those that succeeded [19, 20]. During the outbreak, a panel on the global response to the outbreak noted that 'all of the current approaches to health systems strengthening, from building infrastructure to training professionals, will turn out better and be more cost-effective if data-driven decision-making becomes the norm.' The reports on data and AI-related failures produced at the time of both the Ebola and pandemic outbreaks, and the largely data-driven rapid response to the outbreak as it unfolded, seem to be highlighting much the same analysis: with data-driven systems in general, and AI in particular, some approaches are extraordinarily successful, swiftly shaping the epidemic trajectory. Some manifest other levels and styles of success, providing key data streams and knowledge from the very outset that complement predictions and public health responses; others, as of now, have failed to see real-world adoption. Some, like the outbreak prediction in China, have seen generation one systems surpass existing capabilities and proposals, while systems developed iteratively over time have yet to achieve widespread meaningful policy adoption. Beyond those that failed, even systems with promising predictive potentials have carefully and deliberately been limited either in scope or deployment, to ensure that their use does not outpace their reliability. Finally, some stakeholders are working locally to co-design the AI tools they hope to use and ensure that they will be helpful in practice [21, 22].

Challenges and Future Directions

There are several challenges faced when using artificial intelligence to predict infectious diseases. The first challenge is data quality. Many traditional AI methods for epidemic prediction rely heavily on data quality, but in low- and middle-income countries, syndromic data are affected by various inaccuracies, such as incorrect data entry, fragmentation, missing data, or aggregation errors. The second challenge is the reluctance of many public health practitioners to rely purely on machine learning due to the opacity of the prediction model. The use of a diverse range of data sources to create predictive models for epidemic surveillance and alerting also carries inherent difficulties. Mergeology is a term used to describe the fuzzy nature of integrating multiple data sources to triangulate disease hot spots. Currently, epidemiologists struggle to understand all the mechanisms that lead to disease, and as a result, there is no consensus on one system of models that is strictly better than any other approach [1, 23]. Another limitation of employing AI for epidemic prediction is that algorithms rely heavily on the quality of the data they are trained on. Using historical data to inform models can lead to significant biases, as disease dynamics are constantly evolving, necessitating ongoing retraining for accurate predictions. The integration of big data and AI in public health faces challenges, particularly due to a lack of interdisciplinary collaboration. Effective cooperation among diverse stakeholders is crucial to harness the potential of technological advancements for human welfare. Regulatory frameworks must ensure the protection of individual rights while also facilitating the deployment of new technologies that benefit society, particularly in health, education, and environmental domains. Key challenges include addressing biases and systemic inequities in machine learning algorithms, with much concern over historical data encoding existing biases. This over-reliance on AI may lead to possible discrimination or unfairness. Further discussion is essential to explore solutions and develop tools that make AI-driven disease alert systems viable within infectious disease surveillance. Ultimately, predicting epidemic outbreaks is critical for effective public health preparedness, and while AI holds promise, it faces numerous hurdles that must be addressed to realize its full potential in data-driven response strategies [24, 25].

Ethical Considerations In AI-driven Epidemic Prediction

Data privacy should be a key concern when it comes to developing machine-learning models of disease outbreaks because they rely on cross-referencing data from health-related and non-health-related events to create broad models. This can be especially problematic if the people involved in those events (and their health) did not give informed consent for their information to be used. Further, the existence of explicit and implicit biases in data, as well as unequal access to health care and information, means that the results of AI in public health are likely to have a greater impact on some population subgroups than others. If the data and capabilities of AI systems are not transparent, this may cause even greater harm even if the algorithm has been well-calibrated. Ethical frameworks that allow for the responsible use of AI are available, which we explore, as well as other mechanisms based on human rights law and standards of data governance. Ethical guidelines are at an early stage in development and may not yet reflect the capabilities of AI, meaning that other stakeholders, including members of the public, will need to get involved in discussions about ethical best practices for the use of AI in predicting disease outbreaks. For AI to be a tool that supports public health instead of simply facilitating surveillance, ethical considerations are key. Issues around the use of private and publicly available data, informed consent, and algorithmic bias affecting some populations more than others are all important things to consider. Ethical guidelines are being developed that focus on human rights and equity in AI, but more exploration of these issues with public input is needed, especially as AI and data analytics continue to expand in public health.

The ethical implications are further reaching; however, poor ethical behavior surrounding the implementation of an AI system can destroy stakeholder trust in the entities and government organizations that are meant to be protecting their health, as previous public health surveillance controversies have shown. Ensuring that AI is both ethically developed and used will be key to accurately predicting disease outbreaks in the future [26, 27].

CONCLUSION

The integration of artificial intelligence into epidemic prediction holds immense potential to transform public health strategies. By analyzing large datasets and identifying patterns that traditional methods might miss, AI offers unprecedented capabilities for early outbreak detection and timely intervention. However, significant challenges remain, including issues of data quality, algorithmic bias, and ethical considerations surrounding privacy and equity. Addressing these hurdles requires collaboration among technologists, epidemiologists, policymakers, and communities. With thoughtful implementation and a commitment to ethical practices, AI-driven epidemic prediction can enhance global health systems, save lives, and reduce the socio-economic burden of infectious disease outbreaks. Moving forward, a balanced approach combining technological innovation with human oversight and ethical frameworks will be critical for realizing AI's full potential in epidemic prediction and response.

REFERENCES

1. Zeng D, Cao Z, Neill DB. Artificial intelligence–enabled public health surveillance—from local detection to global epidemic monitoring and control. Artificial intelligence in medicine. 2021. nih.gov
2. Udegbe FC, Nwankwo EI, Igwama GT, Olaboye JA. Real-time data integration in diagnostic devices for predictive modeling of infectious disease outbreaks. Computer Science & IT Research Journal. 2023 Dec;4(3). researchgate.net
3. Khodadadi E, Towfek SK. Internet of Things Enabled Disease Outbreak Detection: A Predictive Modeling System. Journal of Intelligent Systems & Internet of Things. 2023 Nov 1;10(1). [HTML]
4. Keshavamurthy R, Dixon S, Pazdernik KT, Charles LE. Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches. One Health. 2022. sciencedirect.com
5. Majeed A, Hwang SO. Data-driven analytics leveraging artificial intelligence in the era of COVID-19: an insightful review of recent developments. Symmetry. 2021 Dec 23;14(1):16.
6. Kaur I, Behl T, Aleya L, Rahman H, Kumar A, Arora S, Bulbul IJ. Artificial intelligence as a fundamental tool in management of infectious diseases and its current implementation in COVID-19 pandemic. Environmental Science and Pollution Research. 2021 Aug;28(30):40515-32. springer.com
7. Siddique MM, Seraj MM, Adnan MN, Galib SM. Artificial Intelligence for Infectious Disease Detection: Prospects and Challenges. Surveillance, Prevention, and Control of Infectious Diseases: An AI Perspective. 2024 Jul 1:1-22. researchgate.net
8. Chann S, Ly V. AI-Based Clinical Decision Support Systems for Infectious Disease Management and Outbreak Prediction. Asian American Research Letters Journal. 2024. aarlj.com
9. Masum M, Masud MA, Adnan MI, Shahriar H, Kim S. Comparative study of a mathematical epidemic model, statistical modeling, and deep learning for COVID-19 forecasting and management. Socio-Economic Planning Sciences. 2022 Mar 1;80:101249. nih.gov
10. Chiu HY, Hwang CK, Chen SY, Shih FY, Han HC, King CC, Gilbert JR, Fang CC, Oyang YJ. Machine learning for emerging infectious disease field responses. Scientific Reports. 2022 Jan 10;12(1):328. nature.com
11. Coccia M. Sources, diffusion and prediction in COVID-19 pandemic: lessons learned to face next health emergency. AIMS Public Health. 2023;10(1):145.
12. Munagandla VB, Dandyala SS, Vadde BC. AI-powered cloud-based epidemic surveillance system: A framework for early detection. Revista de Inteligencia Artificial en Medicina. 2024 Jul 28;15(1):673-90. redcrevistas.com
13. Ijeh S, Okolo CA, Arowoogun JO, Adeniyi AO, Omotayo O. Predictive modeling for disease outbreaks: a review of data sources and accuracy. International Medical Science Research Journal. 2024 Apr 9;4(4):406-19. fepbl.com
14. Nwankwo EI, Emeihe EV, Ajegbile MD, Olaboye JA, Maha CC. Artificial Intelligence in predictive analytics for epidemic outbreaks in rural populations. International Journal of Life Science Research Archive. 2024 Aug;7(1):078-94. researchgate.net

15. Keshavamurthy R, Charles LE. Predicting Kyasanur forest disease in resource-limited settings using event-based surveillance and transfer learning. Scientific Reports. 2023 Jul 8;13(1):11067.
16. Green M, Pollock FD, Rowe F. New forms of data and new forms of opportunities to monitor and tackle a pandemic. InCOVID-19 and Similar Futures: Pandemic Geographies 2021 Jun 20 (pp. 423-429). Cham: Springer International Publishing. kcl.ac.uk
17. Mhlanga D. The role of artificial intelligence and machine learning amid the COVID-19 pandemic: What lessons are we learning on 4IR and the sustainable development goals. International Journal of Environmental Research and Public Health. 2022 Feb 8;19(3):1879. mdpi.com
18. Rodriguez-Rodriguez I, Rodriguez JV, Shirvanizadeh N, Ortiz A, Pardo-Quiles DJ. Applications of artificial intelligence, machine learning, big data and the internet of things to the COVID-19 pandemic: A scientometric review using text mining. International Journal of Environmental Research and Public Health. 2021 Aug 13;18(16):8578. mdpi.com
19. Mehrdad S, Wang Y, Atashzar SF. Perspective: Wearable Internet of Medical Things for remote tracking of symptoms, prediction of health anomalies, implementation of preventative measures, and …. Frontiers in Robotics and AI. 2021. frontiersin.org
20. Malik YS, Sircar S, Bhat S, Ansari MI, Pande T, Kumar P, Mathapati B, Balasubramanian G, Kaushik R, Natesan S, Ezzikouri S. How artificial intelligence may help the Covid-19 pandemic: Pitfalls and lessons for the future. Reviews in medical virology. 2021 Sep;31(5):1-1. nih.gov
21. Chew AWZ, Zhang L. Data-driven multiscale modelling and analysis of COVID-19 spatiotemporal evolution using explainable AI. Sustainable Cities and Society. 2022. nih.gov
22. Yang H, Zhang S, Liu R, Krall A, Wang Y, Ventura M, Deflitch C. Epidemic informatics and control: A holistic approach from system informatics to epidemic response and risk management in public health. InAI and Analytics for Public Health-Proceedings of the 2020 INFORMS International Conference on Service Science 2021 (pp. 1-46). Berlin/Heidelberg, Germany: Springer. researchgate.net
23. Isgut M, Gloster L, Choi K, Venugopalan J, Wang MD. Systematic review of advanced AI methods for improving healthcare data quality in post COVID-19 Era. IEEE Reviews in Biomedical Engineering. 2022 Oct 21;16:53-69. ieee.org
24. Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, Pirracchio R. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. NPJ digital medicine. 2022 May 31;5(1):66. nature.com
25. Prakash J, Murali L, Manikandan N, Nagaprasad N, Ramaswamy K. RETRACTED ARTICLE: A vehicular network based intelligent transport system for smart cities using machine learning algorithms. Scientific reports. 2024 Nov 6;14(1). [HTML]
26. Xu R, Baracaldo N, Joshi J. Privacy-preserving machine learning: Methods, challenges and directions. arXiv preprint arXiv:2108.04417. 2021. [PDF]
27. Boppiniti ST. Data Ethics in AI: Addressing Challenges in Machine Learning and Data Governance for Responsible Data Science. International Scientific Journal for Research. 2023. isjr.co.in

CITE AS: Kungu Erisa (2024). Artificial Intelligence in Predicting Epidemic Outbreaks. EURASIAN EXPERIMENT JOURNAL OF PUBLIC HEALTH, 7(2):1-6.