



# The Use of AI in Predicting Disease Outbreaks

Kamanzi Ntakirutimana G.

School of Natural and Applied Sciences Kampala International University Uganda

## ABSTRACT

The rapid and unpredictable emergence of infectious diseases continues to pose a significant threat to global health, necessitating more advanced prediction and prevention strategies. The use of Artificial Intelligence (AI) in predicting disease outbreaks has emerged as a powerful tool to navigate the complex biological, environmental, and sociological factors contributing to these outbreaks. AI models, particularly those leveraging machine learning, can analyze vast datasets, detect patterns, and predict the spread of diseases with improved accuracy compared to traditional methods. This paper examines the role of AI in early disease outbreak prediction, current prediction methodologies, and the applications of AI in outbreak forecasting, while also discussing the challenges and limitations associated with AI-driven models.

**Keywords:** Artificial Intelligence, disease outbreaks, machine learning, epidemic prediction, big data analytics.

## INTRODUCTION

The world is a complex and interconnected web of biological systems, socio-cultural dynamics, and physical geography, all of which interact with and affect each other in myriad and complex ways. The ripple effects of changes in any one area can have unforeseen consequences far and wide, such as how the melting of permafrost influences the natural migration patterns of arctic wildlife, or how the introduction of a new agricultural pest species can fundamentally change the locally developed cultivation practices of crops which a native population has grown for centuries. This complexity is what makes understanding disease outbreaks and the factors associated with them and propagation so challenging, as it can be difficult to disentangle the myriad causes and effects involved. A consequence of this complexity is that large datasets often accompany any kind of outbreak event. Using the increasing number of sources for and the ubiquity of many of these datasets, Artificial Intelligence is capable of deriving an understanding of a chaotic system and of making predictions about it. Broadly speaking, this means that outbreaks can be classified as diseases, each with specific biological characteristics and life histories. The specific use of machine learning in modeling disease outbreaks, in this case animal, is then discussed [1, 2]. With artificial intelligence, we can begin to tackle complex systems such as disease outbreaks. Historically, the work of scientists and epidemiologists has focused on developing models to mimic the spread of disease under simple assumptions, such as “X strange animals have been brought to the island” or “the population is divided into groups X, Y, and Z which interact only with each other,” which, while mathematically convenient, lack any actual accuracy. Machine learning models, when fed large enough datasets, make generalizations about a chaotic system without being programmed to do so, in effect learning its laws. In the case of disease outbreaks, the “laws of the system” are the myriad ecological, behavioral, sociological, environmental, and biological factors that influence the propensity of infection, and the virus characteristics associated with it. Given sufficient good data and understanding of these factors, it may be possible to categorize outbreaks as diseases rather than species, each disease with a pattern of outbreak determined by factors associated with it [3].

### **The Importance of Early Disease Outbreak Prediction**

The emergence and rapid spread of infectious diseases pose a significant threat to global public health and safety. Emerging infectious diseases are defined as newly identified strains of pathogens or infectious agents while re-emerging infectious diseases are outbreaks of previously controlled diseases that have resurfaced. Infectious diseases can spread among humans, animals, and plants, with epidemics or pandemics on a national or global scale occurring either in a one-time burst or in a series of waves. Recent examples include the Ebola outbreak in West Africa and the HIV/AIDS pandemic [4]. Prioritizing disease outbreaks in terms of threat to human health, as well as their potential impact, is essential for efficient and effective resource allocation and preventive measure execution. A broad spectrum of pathogens is considered, ranging from the vector-borne viral category encompassing dengue, West Nile, and Zika viruses to the highly pathogenic avian influenza. It is generally assumed that an outbreak will occur only if a certain pathogen is introduced into a specific area with favorable environmental conditions. Decision-makers need to understand the relative threat posed by many pathogens, especially at the beginning of the risk assessment process [5]. The objective is to develop a general risk assessment framework that includes a meta-analysis to review diverse sources of information on prior outbreaks, coupled with database simulations of the spread process and derivatives of methods in machine learning to identify the predictors causing spread from initial known outbreaks. A case study for the dengue epidemic in Florida illustrates the framework's application. While historical outbreaks and expert knowledge play an important role, objective models are crucial to supplement these with the probability that a new outbreak of a specific disease will emerge based on hard data [6, 7].

### **CURRENT METHODS OF DISEASE OUTBREAK PREDICTION**

#### **Traditional and Novel Techniques in The Prediction of Disease Outbreaks**

The growing threat of epidemics and pandemics is having a profound impact on public health and safety, as well as economic stability worldwide. The current COVID-19 crisis brought this important issue to attention. Plans to combat the spread of infectious diseases urgently need to be in place, and the prediction of disease outbreaks can play an important role in this matter. Several techniques are currently used to analyze data that may indicate the emergence of new diseases, and some of these techniques can be complemented with the use of Artificial Intelligence (AI) [8]. The most common techniques currently in use for the surveillance of diseases are built upon the work of the 'Global Early Warning System for Major Animal Diseases' (Global-EWS). This organization was created in response to the H5N1 virus spreading in Southeast Asia in the late 1990s. It was noticed that global bio-surveillance efforts were severely hampered by the lack of access to timely information regarding the emergence of new diseases. This organization was designed to create local and regional Disease Early Warning Systems (DEWS) that can then relay information to a Global DEWS [9]. The principles behind these DEWS are fairly straightforward. Epidemiological data regarding demographic and economic variables of countries, and the incidence of agricultural diseases, are imported from a variety of sources. This data is pre-processed and quality-checked before being analyzed with a set of statistical techniques. These include Bayesian Belief Networks (BBNs) and autoregressive regression models. These statistical techniques produce a variety of outputs, including posterior probabilities of an outbreak occurring, as well as predictions regarding the spread of epidemics across borders. They can provide an effective method of analyzing epidemiological data regarding the emergence of new diseases, though there are several limitations to the use of statistical techniques on their own [10]. In summary, the well-established DEWS has been created by the FAO/IAEA International Atomic Energy Agency. These DEWS can perform the analysis of epidemiological data using a variety of statistical techniques. However, these systems have several limitations that need to be considered, including publication bias present in epidemiological data, the time lag between an outbreak and the publication of disease reports that may inhibit the detection of outbreaks, and the neglect of the spatial component of epidemiological data. Data must be analyzed with techniques that can account for these factors [11].

#### **Applications of AI in Disease Outbreak Prediction**

The rapid advancement of Artificial Intelligence (AI) and machine learning are significant developments that are transforming a variety of disciplines, including healthcare, agriculture, climate science, and epidemic prediction. Infectious diseases, which are brought on by viruses or microbes and spread directly or indirectly from one organism to another, continue to threaten public health across the globe, particularly in developing countries. In recent years, there has been a flurry of research exploring AI and machine learning approaches to predict the emergence of new diseases based on big data, and rapidly

spreading old diseases. These studies often rely on similar indicators to evaluate past outbreak data, such as environmental, ecological, demographic, and socio-economic factors [12]. There are two main categories of AI applications for disease outbreak prediction. The first is based on environmental and ecological time series models, which focus on climate and environmental aspects such as temperature, rainfall, humidity, vegetation, and water bodies. These have been used to predict Cholera, Dengue, Malaria, Rift Valley Fever, and Zika. The second category is based on mining large-scale spatio-temporal epidemiological event datasets, particularly on social media such as Twitter. Data analytics is used to create models to detect noise, early warning signs, spatio-temporal spreading behavior, and epidemic risk perception of various diseases including the flu, Dengue, and food poisoning. Despite the increasing popularity of using big data spatio-temporal analysis to better monitor and manage disease outbreaks in near real-time, most attempts are still limited to more developed countries [13]. In a highly connected world with unprecedented movement of people and goods, the emergence and spread of newly evolved infectious diseases pose a major threat to global health and the economy. A set of data-driven models of rapid spatio-temporal spreading behavior has been proposed based on large-scale historical epidemic datasets of infectious diseases that have infected humans. Epidemiological spreading models have been used to describe the spatio-temporal dynamics of various diseases on simple networks, where the spreading can exhibit wave-like propagation or irregular diffusion patterns depending on the type of infection. To incorporate the heterogeneity of both the spreading process and the geographical connection, approaches have been proposed to represent a disease as a spatiotemporal field on top of the realistic transportation network between different locations (e.g., airline flight network). Such models have been used to simulate the international spread of diseases by using aggregated datasets of the networks between airports. The accuracy of this model has been successfully validated against past cholera epidemics in the Haiti slums, the spread of the H1N1 virus, the Ebola outbreak, and the onset of the Zika virus across the Americas [14, 15].

#### Challenges and Limitations of AI in Disease Outbreak Prediction

While AI has shown great promise in predicting disease outbreaks, it also presents several challenges and limitations. A major challenge is the issue of data availability and quality. AI systems rely on large datasets for training and validation, and the lack of high-quality, comprehensive data can hinder the effectiveness of these systems. For instance, certain diseases may have limited historical data, making it difficult to build accurate predictive models. Additionally, the data quality can also vary, with some datasets having inconsistencies or biases that can impact the model's performance [16]. Another challenge is the interpretability and explainability of AI models. Many AI algorithms, particularly deep learning models, are often considered "black boxes" as they can be complex and difficult to understand. This lack of transparency can make it challenging for public health officials and policymakers to trust and effectively utilize AI predictions. There may also be ethical concerns regarding the use of AI in this context, such as issues around data privacy and the potential for bias in the algorithms [17]. Finally, it is essential to recognize that AI is not a panacea for disease outbreak prediction. AI models can complement traditional epidemiological approaches, but they should not be viewed as a substitute. Therefore, collaboration between data scientists and epidemiologists is necessary to develop effective AI solutions for predicting disease outbreaks [18].

#### CONCLUSION

AI can significantly improve disease outbreak prediction and provide timely insights for public health treatments. AI can detect new dangers more effectively than traditional statistical methods by using vast datasets and constructing prediction models that account for a variety of environmental, biological, and societal elements. However, major impediments continue to exist, including data quality, model interpretability, and ethical considerations. A multidisciplinary approach that combines AI with traditional epidemiological methodologies is required to maximize disease outbreak prediction and ensure a strong public health response.

#### REFERENCES

1. Hamzah FB, Lau C, Nazri H, Ligot DV, Lee G, Tan CL, Shaib MK, Zaidon UH, Abdullah AB, Chung MH. CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. Bull World Health Organ. 2020 Mar 19;1(32):1-32. [archive.org](https://www.who.int/publications/iitem/9789240000000)
2. Shuja J, Alanazi E, Alasmay W, Alashaikh A. COVID-19 open-source data sets: a comprehensive survey. Applied Intelligence. 2021 Mar;51(3):1296-325.

3. Heidari A, Jafari Navimipour N, Unal M, Toumaj S. Machine learning applications for COVID-19 outbreak management. *Neural Computing and Applications*. 2022 Sep;34(18):15313-48. [springer.com](https://www.springer.com)
4. Zumla A, Niederman MS. The explosive epidemic outbreak of novel coronavirus disease 2019 (COVID-19) and the persistent threat of respiratory tract infectious diseases to global health security. *Current opinion in pulmonary medicine*. 2020 May 1;26(3):193-6.
5. Lal A, Ashworth HC, Dada S, Hoemeke L, Tambo E. Optimizing pandemic preparedness and response through health information systems: lessons learned from Ebola to COVID-19. *Disaster medicine and public health preparedness*. 2022 Feb;16(1):333-40. [cambridge.org](https://www.cambridge.org)
6. Ioannidis JP, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. *International journal of forecasting*. 2022 Apr 1;38(2):423-38.
7. Jia L, Li K, Jiang Y, Guo X. Prediction and analysis of coronavirus disease 2019. *arXiv preprint arXiv:2003.05447*. 2020 Mar 11.
8. Coccia M. Sources, diffusion, and prediction in COVID-19 pandemic: lessons learned to face next health emergency. *AIMS Public Health*. 2023;10(1):145.
9. Imperia E, Bazzani L, Scarpa F, Borsetti A, Petrosillo N, Giovanetti M, Ciccozzi M. Avian influenza: could the H5N1 virus be a potential next threat? *Microbiology Research*. 2023 May 9;14(2):635-45. [mdpi.com](https://www.mdpi.com)
10. Xu B, Gutierrez B, Mekar S, Sewalk K, Goodwin L, Loskill A, Cohn EL, Hswen Y, Hill SC, Cobo MM, Zarebski AE. Epidemiological data from the COVID-19 outbreak, real-time case information. *Scientific data*. 2020 Mar 24;7(1):106. [nature.com](https://www.nature.com)
11. Becker SM, Franz TE, Morris TC, Mullins B. Field Testing of Gamma-Spectroscopy Method for Soil Water Content Estimation in an Agricultural Field. *Sensors*. 2024 Mar 30;24(7):2223.
12. . Khanday AM, Rabani ST, Khan QR, Rouf N, Mohi Ud Din M. Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*. 2020 Sep; 12:731-9. [springer.com](https://www.springer.com)
13. Pham QV, Nguyen DC, Huynh-The T, Hwang WJ, Pathirana PN. Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts. *IEEE access*. 2020 Jul 15; 8:130820-39. [ieee.org](https://www.ieee.org)
14. Chew AW, Zhang L. Data-driven multiscale modelling and analysis of COVID-19 spatiotemporal evolution using explainable AI. *Sustainable Cities and Society*. 2022 May 1; 80:103772.
15. Marsland III R, Mehta P. Data-driven modeling reveals a universal dynamic underlying the COVID-19 pandemic under social distancing. *MedRxiv*. 2020 Apr 24.
16. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, Gallivanone F, Cozzi A, D'Amico NC, Sardanelli F. AI applications to medical images: From machine learning to deep learning. *Physica medica*. 2021 Mar 1; 83:9-24. [physicamedica.com](https://www.physicamedica.com)
17. Chou YL, Moreira C, Bruza P, Ouyang C, Jorge J. Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications. *Information Fusion*. 2022 May 1; 81:59-83.
18. 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](https://www.nih.gov)

**CITE AS: Kamanzi Ntakirutimana G. (2024). The Use of AI in Predicting Disease Outbreaks. *Research Output Journal of Biological and Applied Science* 4(2):19-22. <https://doi.org/10.59298/ROJBAS/2024/421922>**