IDENTIFYING THE APPROPRIATE PREDICTION METHOD FOR DISEASE OUTBREAK SURVEILLANCE AND FORECASTING

A Narrative Analysis of Existing AI Models

A Project Report

Student: Izz

Mentor: Dr Hana Antonio

Sapiente Education

November, 2026

1. Introduction

In December 2019, Wuhan became the origin of the COVID-19 disease that became a global pandemic and eventually an ongoing endemic in May 2023. (1,2) However, it was not until December 2020 that vaccines began distribution to the public, at which point the cumulative global death toll from COVID-19 was over 1.8 million. (1,2) In addition, there were limited testing kits for disease surveillance, leaving many infected people untreated, and COVID-tracking methods only began public distribution in mid to late 2020. (1,2)

Digital tracing is one example of technology advancements in public health. Nowadays, Artificial Intelligence (AI) has become an important tool for humans' daily activities. (3,4) In health, the example application of AI is assisting in analysing image scans to detect abnormalities in patients. (3,4) AI can also be useful in predicting disease outbreaks by helping to detect infectious microbes earlier for countermeasures to be produced more quickly. (3,4) However, evaluation of its performance needs to be studied further.

2. Aim

This article is aimed to evaluate the performance of AI models in disease surveillance and outbreak forecasting/prediction through rapid literature review.

3. Method

We decided to do a simple literature review by searching available studies on PubMed from the last five years. We limited the studies that were written in English language. Data will be extracted into tabular format with information about the study title, type of AI tools used, type of data used, and summary of the paper about AI role in infectious disease. Narrative analysis will be done based on the findings.

4. Results

No	Paper Title	Link	Type of AI Used	Data Source/ Type	Key Finding
1	Artificial intelligence in early warning systems for infectious disease surveillance: a systematic review	https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2025.1609615/full	Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP)	Epidemiological, web, climate, wastewater data	Benefit: Earlier outbreak detection and improved prediction accuracy. Limitation: Data quality and bias model transparency, system

					integration difficulties, privacy and equity concern.
2	A reproducible ensemble machine learning approach to forecast dengue outbreaks	https://www.nature.com/articles/s41598-024-52796-9	ML – ensemble models such as CatBoost, Support Vector Machine (SVM), Long Short Term Memeory (LSTM)	Dengue Incidence Rate (DIR) data, eco-climatic, environmental and population factors data in Brazil.	Benefit: Can predict DIR trends with low uncertainty. Limitation: Struggles with extreme values.
3	Detection of COVID-19 epidemic outbreak using machine learning	https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1252357/full	ML - using classifiers such as SVM, Random Forest (RF), XGBoost	Epidemiological COVID-19 data (Korea)	High accuracy in estimating classification of transmission trends and successful prediction of start of new outbreaks. RF and XGBoost showed the highest accuracy.
4	Explainable artificial intelligence for predicting dengue outbreaks in Bangladesh using eco-climatic triggers	https://pubmed.ncbi.nlm.nih.gov/40548126/	ML – XGBoost, LightGBM	Epidemiological dengue data, Dengue deaths and cases, climate and socio-demographic data and landscape (Bangladesh)	XGBoost excelled in predicting dengue outbreaks through population density, precipitation, temperature, and land-use patterns.

5	A comprehensive review of artificial intelligence in prevention and treatment of COVID-19 pandemic	https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.845305/full	ML - Time series analysis with ARIMA, and SARIMA approaches	Epidemiological COVID-19 data	Prediction models are only a way to provide an early warning. Combination of mathematical modelling and machine learning can be a future development of AI in infectious disease outbreaks.
6	Artificial intelligence in public health: the potential of epidemic early warning systems	https://journals.sagepub.com/doi/pdf/10.1177/03000605231159335	DL – NLP (BERT transformer architecture)	Open-source epidemiological data (Past outbreaks, travel, environmental factors, vector distribution, satellite meteorological data)	Can enable early intervention early in the timeline of an epidemic and allow newly emergent epidemics to be identified. The system has a large unfiltered data, but the uptakes by the public authorities are low.
7	Systematic determinants of global COVID-19 burden: longitudinal time-series analysis using big data-driven AI	https://www.jmir.org/2025/1/e79745	ML – time series analysis, gradient boosting, and XGBoost models	Global COVID-19 data (38 countries)	Time series analysis model using COVID-19 big data can help to generate metrics for future COVID - 19 pandemic preparedness that combines viral evolution, immunity dynamics, and environmental factors.

5. Discussion

Machine learning (ML) is the capacity of systems to learn from problem-specific training data to automate analytical model building and solve associated tasks. (5,10) ML algorithms are categorised into supervised learning, unsupervised learning, time series analysis, and reinforcement learning. (5,10) Supervised learning trains on labeled data to predict outcomes for new, unseen inputs to predict the likelihood and timing of future outbreaks. (5,10) Ensemble learning is one of the robust supervised ML methods for outbreaks forecasting. (5,10) Boosting algorithms such as XG Boost, CatBoost, LightBoost are commonly used as they reduce risk of variance and bias and lead to better predictions. (5,10) Unsupervised learning trains on unlabeled data to find hidden structures, anomalies or groupings without predefined outcomes, which allows for identifying clustering cases or outbreaks. (5,10) The clustering method is one of the most common methods used to uncover transmission dynamics and identify anomaly cases. (5,10)

Time series analysis is another type of ML that can be used to predict disease incidence and trend based on historical data patterns. (5,10) The most common approach is Seasonal Autoregressive Integrated Moving Average Model (SARIMA) - where you can predict seasonality,trends, and fluctuation of a disease, such as influenza. (5,10)

Deep learning (DL) is a machine learning concept based on artificial neural networks. DL improves forecasting outbreaks by evaluating large datasets and recognising complex patterns. (5,10) NLP is one of the common DL techniques used in disease outbreaks. (5,10) NLP allows extracting relevant information from large amounts of online textual data and improves early outbreak detection through identifying mentions of symptoms and self-reported illnesses. (5,10) In terms of architectures, BERT transformers are commonly used, as they can classify and prioritise textual information for faster and more efficient identification of outbreak-relevant data. (5,10)

A deeper application of AI in infectious disease can be seen from the recent COVID- 19 pandemic. (5, 7-11)During the pandemic, platforms like BlueDot provided early detection of the outbreak, while AI analysed social media and search query data to track the virus's spread and monitor public sentiment. (5, 7-11) Other direct applications of AI in infectious disease outbreaks can be seen in dengue and influenza. (5,6) Dengue research about AI focuses on forecasting cases or outbreaks through epidemiological surveillance data combined with climate or meteorological variables. (5,6) In addition, AI is being used to design dengue early warning systems to provide timely predictions and support public health interventions and vector control efforts. (5,6) As for influenza, Google Flu Trends was an early attempt to predict disease activity using search queries while EPIWATCH is used for predicting flu season severity. (5) DL models have been developed for influenza surveillance using influenza-like illness (ILI) data. (5) This rapid review only searched for limited literature, hence, it only represented application of several infectious diseases such as COVID-19, influenza, and dengue fever. More studies need to be done in application of AI in infectious disease outbreaks.

6. Conclusion

In summary, AI use in disease surveillance and outbreak prediction can help in disease containment and prevent further transmissions, which can ensure that epidemics and pandemics become less severe and widespread in the future. However, as AI still possesses limitations in how it analyses data, the results produced may not always be accurate. Real world data will be needed to make prediction and surveillance more accurate.

References

1. Asher S. TraceTogether: Singapore turns to wearable contact-tracing Covid tech. BBC News [Internet]. 2020 Jul 5 [cited 2026 Nov 10]. Available from: https://www.bbc.co.uk/news/technology-53146360

2. Centers for Disease Control and Prevention. CDC Museum COVID-19 timeline [Internet]. Atlanta (GA): CDC; [cited 2026 Nov 10]. Available from: https://www.cdc.gov/museum/timeline/covid19.html

3. Zhao AP, Zhao A, Li S, Cao ZD, Hu P, Wang J, et al. AI for science: predicting infectious diseases. J Saf Sci Resil. 2024;5(2):130-146.

4. Parums DV. Editorial: infectious disease surveillance using artificial intelligence (AI) and its role in epidemic and pandemic preparedness. Med Sci Monit. 2023;29:e941209.

5. Villanueva-Miranda I, Xiao G, Xie Y. Artificial intelligence in early warning systems for infectious disease surveillance: a systematic review. Front Public Health. 2025;13:1609615.

6. Sebastianelli A, Spiller D, Carmo R, Wheeler J, Nowakowski A, Jacobson LV, et al. A reproducible ensemble machine learning approach to forecast dengue outbreaks. Sci Rep. 2024;14(1):3807.

7. Cho G, Park JR, Choi Y, Ahn H, Lee H. Detection of COVID-19 epidemic outbreak using machine learning. Front Public Health. 2023;11:1252357.

8. Rahman MS, Shiddik MAB. Explainable artificial intelligence for predicting dengue outbreaks in Bangladesh using eco-climatic triggers. Glob Epidemiol. 2025;10:100210.

9. Wang H, Jia S, Zhao L. A comprehensive review of artificial intelligence in prevention and treatment of COVID-19 pandemic. Front Genet. 2022;13:845305.

10. MacIntyre CR, Chen X, Kunasekaran M, Quigley A, Lim S, Stone H, et al. Artificial intelligence in public health: the potential of epidemic early warning systems. J Int Med Res. 2023;51(3):3000605231159335.

11. Cao Z, Han W, Zhang X, Zhang C, Zeng J, Chen Y, et al. Systematic determinants of global COVID-19 burden: longitudinal time-series analysis using big data-driven artificial intelligence. J Med Internet Res. 2025;27:e79745.