Journal of Medical Science, Biology, and Chemistry (JMSBC)
ISSN: 3079-2576 (Online)
Volume 2 Issue 1, (2025)
https://doi.org/10.69739/jmsbc.v2i1.465
https://journals.stecab.com/jmsbc

Published by
Stecab Publishing

Research Article

AI and Machine Learning for Early Detection of Infectious Diseases in the US: Opportunities and Challenges

[1]Adekunle Adeoye, [2]Chibuzo Okechukwu Onah, [3]Enibokun Theresa Orobator, [4]Elizabeth Anuoluwa Akintayo, [5]Jacob Ukpabio Inuaeyen, [6]Daniel Ojonugwa Umoru, [7]Isaiah Olumeko, [8]Gladys Jacob Dick, [9]Kelechi Wisdom Elechi, *[10]Ubalaeze Solomon Elechi, [11]Muhammad Bello Demola

About Article

Article History

Submission: March 19, 2025
Acceptance : April 26, 2025
Publication : May 03, 2025

Keywords

AI-Driven Disease Surveillance, Computational Epidemiology, Health Informatics, Smart Public Health Systems, Digital Epidemiology Tools

About Author

[1] Department of Mathematics and Statistics, Georgia State University, 33 Gilmer Street SE, Atlanta, GA 30303, USA

[2] J. Mack Robinson College of Business, Georgia State University, 33 Gilmer Street SE, Atlanta, GA 30303, USA

[3] College of Medicine and Veterinary Medicine, University of Edinburgh, UK

[4] Department of Bioengineering and Biomedical Engineering, Wayne State University, 45 Warren Ave, Detroit, MI, USA

[5] Department of Pharmaceutical and Medicinal Chemistry, University of Nigeria, Enugu, Nigeria

[6] Department of Pharmacy, National Hospital Abuja, Nigeria

[7] College of Pharmacy, University of Houston, Texas, USA

[8] Elmleigh Hospital, Havant, UK

[9] School of Biomedical Sciences, University of Texas Health Science Center, San Antonio, USA

[10] Department of Radiography and Radiological Sciences, University of Nigeria, Enugu Campus, Nigeria

[11] Department of Computer Science and Technology, Ulster University, Birmingham, UK

Contact @ Ubalaeze Solomon Elechi
elechiuba@gmail.com

ABSTRACT

Artificial intelligence and machine learning are reshaping U.S. infectious-disease surveillance by rapidly integrating clinical, environmental, and open-source data to flag anomalies sooner than conventional methods. This article aims to assess how artificial intelligence (AI) and machine learning (ML) can accelerate early detection of emerging infectious diseases in the United States In case studies, real-time ML tools cut hospital-acquired infections by 40%, and systems like BlueDot predicted major outbreaks days before official alerts, underscoring strong gains in speed and accuracy. However, biased training data, opaque "black-box" models, privacy risks, and high implementation costs still threaten equitable, trustworthy deployment. Overcoming these barriers will require rigorous data-quality standards, explainable algorithms, interdisciplinary governance, and scalable validation frameworks—advances that could extend early-warning capacity from local hospitals to global health security networks.

Citation Style:
Adeoye, A., Onah, C. O., Orobator, E. T., Akintayo, E. A., Inuaeyen, J. U., Umoru, D. O., Olumeko, I., Dick, G. J., Elechi, K. W., Elechi, U. S., & Demola, M. B. (2025). AI and Machine Learning for Early Detection of Infectious Diseases in the US: Opportunities and Challenges. *Journal of Medical Science, Biology, and Chemistry, 2*(1), 54-63. https://doi.org/10.69739/jmsbc.v2i1.465

1. INTRODUCTION

Infectious diseases have emerged and re-emerged throughout human history, a dynamic that continues to influence public health perspectives and societal structures (CDC, 1994). Several noteworthy emerging infectious disease events have occurred in the United States, including foodborne outbreaks of Escherichia coli O157:H7, the rise of drug-resistant pneumococcal infections, waterborne illnesses caused by Cryptosporidium, and the more recently identified hantavirus pulmonary syndrome (CDC, 1994). Concerningly, the frequency of these occurrences has been on the rise, with new cases of infectious diseases associated with wildlife zoonosis rising steadily every ten years since 1940. Human activities that disturb ecological balances and invade natural ecosystems are mostly to blame for this increase, leading to increased human-animal contact and the potential for disease reservoirs (Emerging Infectious Disease, 2025). These emerging infections constitute a significant portion of the overall infectious disease landscape, accounting for at least 15% of all human pathogens (McArthur, 2019). Compared to the average from 2017 to 2019, the number of reported Lyme disease cases in the United States skyrocketed to 62,000 in 2022, an almost 70% rise. This significant increase may be linked to various factors, such as modifications in surveillance definitions and the broader geographical distribution of tick vectors (CIDRAP, 2024). In addition, by early December 2024, the US saw a notable increase in measles cases, with more than 280 reported, representing the highest annual total in the past five years.

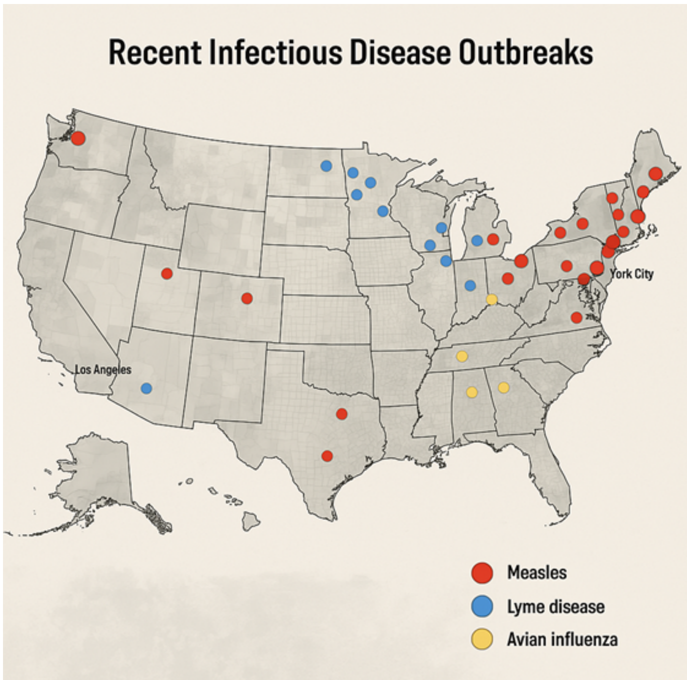

Figure 1. Geographic distribution of recent infectious disease outbreaks in the United States, highlighting regions with increased measles and Lyme disease incidence

Public health is significantly impacted by infectious diseases because they increase rates of morbidity and mortality. In 2023, infectious diseases claimed almost three million lives in the US, even though COVID-19, the top cause at the time, fell to tenth place (Ahmad, 2024). In addition to the direct health impacts, the economic burden of infectious diseases is substantial. Estimates indicate that the direct and indirect costs, encompassing healthcare expenditures, economic losses, and days of disability, may surpass $120 billion each year (CDC, 1994). The COVID-19 pandemic highlighted this issue, with the overall economic damage to the US projected to reach approximately $14 trillion by the conclusion of 2023 (CDC Foundation, 2024). The confluence of persistent hazards, increasing incidence, and significant economic consequences underscores the pressing need for innovative approaches to detect infectious diseases in their early stages.

Prompt identification of emerging infectious illnesses is crucial for the implementation of rapid and effective public health measures. Control measures such as patient isolation, rigorous contact tracing, and quarantine protocols can be promptly enacted upon the early detection of an outbreak. An outbreak's course can be changed, and its potential to develop into a worldwide pandemic or widespread epidemic can be prevented by engaging in these measures (MacIntyre *et al.*, 2023). Early warning systems also offer critical time for the development and implementation of medical countermeasures, such as vaccines, treatments, and diagnostics, as well as for the execution of more comprehensive public health containment and mitigation plans (Zhou *et al.*, 2024). The capacity to identify and react promptly to emerging risks offers the potential to reduce related morbidity, death, and significant economic disruptions that may arise from large-scale outbreaks (Chae *et al.*, 2018).

Artificial intelligence and machine learning have become formidable instruments in data analysis, possessing the capacity to transform early detection initiatives. These technologies are proficient at processing and analyzing extensive volumes of intricate data from various sources, detecting nuanced patterns and abnormalities that may indicate the onset of an infectious disease outbreak (MSc, 2024). Early identification of pathogen spread, less strain on human resources for monitoring, and the creation of better-informed recommendations for preventative public health actions could all be attainable by combining AI and ML capabilities with conventional surveillance techniques (MSc, 2024). The objective of this review is to present a thorough and critical examination of the state of AI and machine learning applications in the early identification of newly emerging infectious illnesses in the US. The several AI and ML techniques that are pertinent to this task will be examined, together with their possible advantages and opportunities, as well as the difficulties and restrictions that come with putting them into practice. The review will analyze specific case studies of AI-driven disease surveillance systems in use or under development in the US, providing comparative insights and emphasizing lessons gained. This review shows that embedding AI and ML in U.S. disease-surveillance pipelines is indispensable for catching pathogens at their first signal, containing outbreaks before they spread, and steering future research toward transparent, equitable, and scalable early-warning systems.

2. LITERATURE REVIEW

2.1. Overview of AI and Machine Learning in Epidemiology

Epidemiology is changing in many ways due to the use of AI and machine learning techniques, especially when it comes to early detection and surveillance of infectious diseases. Numerous fundamental strategies are pivotal to this transition. Using historical data, statistical algorithms, and machine learning, predictive analytics is essential for predicting future health outcomes and spotting possible concerns before they become serious problems. These models are capable of estimating the risk of developing particular diseases in people or groups by examining a wide range of factors, such as lifestyle choices, genetic predispositions, environmental factors, and past health events (SCIP, 2024). Predictive analytics can identify people who are more vulnerable to infectious diseases, allowing for more focused preventative actions. Furthermore, these models are crucial for the early identification of epidemics by monitoring trends in patient symptoms, environmental variables, and comprehensive population health data (SCIP, 2024). The COVID-19 pandemic underscored the efficacy of these models in resource distribution and containment tactics (SCIP, 2024). Diverse machine learning algorithms, such as logistic regression, random forests, support vector machines, and gradient boosting machines, are utilized in epidemiological applications for categorizing the presence or absence of an illness based on clinical data (Epelde, 2024). The capacity of predictive analytics to utilize patterns in varied datasets provides a substantial benefit in the proactive management of infectious disease risks. A real-time machine learning system successfully prevented up to 40% of nosocomial infections.

Natural language processing provides a robust method for early detection through the study of unstructured data sources. NLP approaches facilitate the identification and extraction of pertinent information from extensive text archives, including clinical notes, news articles, social media posts, and other publicly available data (MacIntyre *et al.*, 2023). This capability is particularly valuable for infectious disease surveillance, as early signals of outbreaks may first appear in informal channels before being officially reported. For instance, NLP has been used to accurately detect urinary tract infections from hospital records with high sensitivity (87% to 100%) and specificity (94% to 100%) (Omar *et al.*, 2024) and to conduct surveillance for Lyme disease by analyzing social media activity with 90% accuracy using a BERTweet model (Westat, n.d.). Moreover, NLP can improve the comprehensiveness of immunization data by extracting information from the narrative components of electronic health records (Westat, n.d.). The process of pathogen and outbreak detection utilizing NLP generally encompasses four fundamental components: analyzing open-source data to detect early warning indicators, discerning regional patterns within these signals, modeling outbreak behavior, and promptly identifying misinformation that may obstruct public health initiatives (MSc, 2024). NLP leverages the extensive information in unstructured text, thus enhancing epidemiological surveillance and enabling earlier identification of possible health emergencies.

Deep learning, a sophisticated subset of machine learning, has exhibited exceptional proficiency in evaluating intricate and high-dimensional data, which is increasingly common in epidemiological research (Zhou *et al.*, 2024). Deep learning models, including deep neural networks and long short-term memory networks, have demonstrated enhanced efficacy in forecasting infectious disease trends relative to conventional statistical methods (Chae *et al.*, 2018). These models excel at identifying intricate, non-linear relationships within data, making them particularly well-suited for tasks like image analysis and feature extraction. In the area of infectious diseases, deep learning has been successfully applied to detect pneumonia in chest X-rays with high accuracy (up to 99% specificity) (Melchane *et al.*, 2024), identify malaria parasites in blood samples (Aiforia, 2021), and categorize bacterial strains with high accuracy (Malesu, 2025). Moreover, deep learning holds promise for identifying early signals of disease transmission from internet data, detecting pathogens in environmental samples, and predicting the risk of zoonotic spillover events (Zhou *et al.*, 2024). Deep learning's sophisticated analytical capabilities offer essential tools for augmenting diagnostic precision and refining predictive models in combating emerging infectious illnesses.

3. METHODOLOGY

This evaluation prioritized breadth over systematic rigor by using an organized yet adaptable technique to synthesize the body of information on AI/ML applications for emerging infectious disease identification in the United States. In accordance with comprehensive review standards, the methodology concentrated on three stages: thematic synthesis, iterative literature exploration, and scope specification.

3.1. Scope and Search Strategy

Using data from PubMed, Scopus, IEEE Xplore, and Google Scholar, the analysis included case studies, peer-reviewed research, and gray literature published between 2018 and 2025. Boolean operators were used to refine the search results, which included "AI in disease surveillance," "outbreak prediction using machine learning," and "infectious disease detection in the US." As new themes emerged, the search technique changed repeatedly, adding additional terms (such as "deep learning zoonosis").

We used the following criteria to determine which studies we included:

1. Addressed U.S. public health contexts or transferable methodologies;
2. Offered empirical insights, theoretical frameworks, or real-world implementations; or
3. Concentrated on AI/ML applications for infectious disease surveillance or early detection.

Only diagnostic studies, non-peer-reviewed articles, and publications written in languages other than English were excluded.

3.2. Synthesis and Analysis of Data

Findings were divided into five categories using a hybrid theme analysis approach: case studies (e.g., CDC's Data Modernization Initiative, BlueDot), deep learning applications, natural language processing, predictive analytics, and implementation problems. Technology developments were contextualized by cross-referencing with public health data and outbreak timings (such as COVID-19 and Lyme disease spikes).

3.3. Quality Considerations

Journal impact indicators, institutional affiliations, and methodological transparency were used to evaluate the legitimacy of the sources. Case studies that were referenced in several reliable sources were given preference. Without being constrained by formal methods, this methodology allowed for a critical assessment of AI/ML's developing involvement in infectious disease surveillance in the United States by striking a balance between comprehensiveness and narrative coherence.

4. RESULTS & DISCUSSION

4.1. Opportunities and Potential for Early Detection

The incorporation of artificial intelligence and machine learning into epidemiological practices provides a wealth of opportunities for improving the early detection of emerging infectious diseases, significantly surpassing traditional surveillance methods (Srivastava *et al.*, 2025).

The ability of AI and ML to conduct swift data analysis is one of their most significant benefits (Brownstein *et al.*, 2008). The pace at which these technologies can handle large amounts of data from many sources, such as social media, news outlets, electronic health records, environmental monitoring systems, and travel records, is significantly faster than that of humans (Suvvari & Kandi, 2024). For instance, HealthMap, an AI-driven system, utilizes a continuously updated dictionary to extract geographical information from online reports in multiple languages, enabling the identification of potential disease clusters across the globe (HealthMap, 2024). In a similar vein, the Centers for Disease Control and Prevention are utilizing AI and ML to examine extensive and intricate datasets, which include images, audio, unstructured text, and genomic data, to extract meaningful insights and guide public health initiatives (CDC, 2025). The capacity to rapidly analyze and integrate various data streams is essential in the realm of emerging infectious diseases, where initial indicators may be faint and dispersed among multiple sources. The swift analytical abilities of AI and ML enable the prompt recognition of potential threats that could otherwise go unnoticed or be detected only after considerable escalation has taken place. AI has demonstrated a significant reduction in diagnostic time in radiology and pathology, achieving approximately 90% or more efficiency (Jeong *et al.*, 2025). BlueDot has announced a reduction in manual detection time by almost 90% with its advanced solution (Cade, 2024).

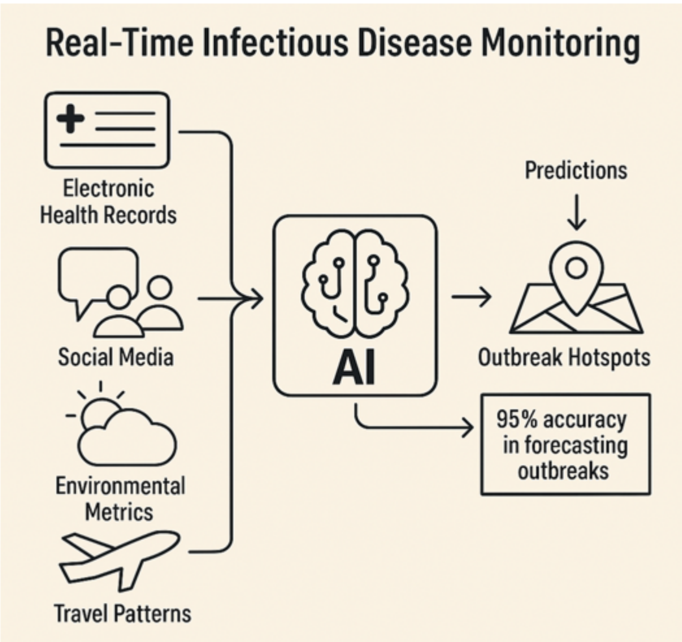

Figure 2. This diagram showcases an AI-powered surveillance system integrating real-time data streams from EHRs, social media platforms, and environmental monitoring systems. Such systems have demonstrated predictive accuracies of up to 95%, significantly reducing manual detection time by nearly 90%, as seen in platforms like BlueDot.

Additionally, machine learning algorithms have shown the capability to greatly enhance the predictive accuracy of forecasting disease outbreaks (Center for Global Digital Health Innovation, n.d.). Through the examination of historical data and the identification of intricate patterns and correlations, machine learning models are capable of forecasting disease outbreaks with a precision that frequently surpasses that of conventional statistical approaches (Melchane *et al.*, 2024). An exemplary case is BlueDot, an AI-driven platform that effectively detected the onset of the COVID-19 outbreak several days before the World Health Organization's official alert (BlueDot Leverages Data Integration to Predict COVID-19 Spread, n.d.). Furthermore, AI models have demonstrated the capability to accurately predict influenza trends, which supports effective preparedness and resource allocation (Ekundayo, 2024). Research has indicated impressive accuracy levels in particular applications, with a BERTweet model reaching 90% accuracy in identifying Lyme disease from social media data and a neural network-based program nearing 96% accuracy in influenza screening in China (McCormick, 2025; Malesu, 2025). The achievements highlight the significant role of AI and ML in improving our capacity to predict and respond to infectious disease outbreaks, facilitating more precise and efficient public health interventions. A method for predicting AI-related risks identified 1.85 times more asymptomatic, infected travelers than random surveillance testing (Ghaffar Nia *et al.*, 2023).

The real-time monitoring of infectious disease activities is another significant benefit of AI and ML in this field. These systems can continually monitor many data streams to identify early indicators of disease development and anticipate possible outbreaks as they develop (Zhou *et al.*, 2024). This ability is demonstrated by AI-driven early warning systems like HealthMap, which analyze real-time data from numerous web-based sources to quickly identify conversations and reports of new infections. AI-enabled mobile health data integration improves real-time monitoring even more by enabling continuous tracking of population movements and individual health indicators, which makes it easier to identify and react quickly to possible health hazards (Janet Aderonke Olaboye *et al.*, 2024). Public health officials can take preventative steps and slow the spread of newly developing infectious diseases because of their ongoing alertness and capacity to process information as it becomes available

4.2. Challenges and Limitations in Implementation

When applied in actual infectious disease monitoring settings, artificial intelligence (AI) and machine learning systems encounter considerable challenges despite their apparent advantages (Ali *et al.*, 2025). Our research has found some significant obstacles that hinder their effective implementation. Data bias emerges as a significant obstacle we have faced. When AI and ML models are trained on datasets that do not accurately reflect America's ethnic diversity, the repercussions can be significant; computers effectively adopt these biases, resulting in detection and response gaps that replicate current healthcare inequalities (Parums, 2023). A model primarily trained on data from a single racial group typically encounters difficulties when faced with patients from diverse backgrounds (Accuray, n.d.). This limitation isn't merely theoretical. Obermeyer's team uncovered commercial algorithms that, by using healthcare spending as a proxy for illness, substantially underestimated Black patients' care needs compared to White patients with identical conditions (Dankwa-Mullan, 2024). We've observed two primary mechanisms behind these disparities: explicit exclusion (where certain communities remain systematically absent from datasets) and environmental bias (where social determinants of health disproportionately impact specific populations) (Dankwa-Mullan, 2024). If we do not address these data biases directly, our early detection systems will unavoidably sustain, rather than alleviate, inequitable health outcomes among various American populations.

The "black box" issue constitutes a significant obstacle. Numerous high-performing AI systems, particularly those constructed on intricate deep learning frameworks, function with a level of opacity that perplexes even their creators (Ennab & Mcheick, 2024). Clinicians and public health officials justifiably exhibit skepticism toward algorithms when they cannot comprehend the rationale behind essential recommendations (Epelde, 2024). A recent CDC survey indicates that this interpretability gap continues to be a significant obstacle to adoption among health departments countrywide. (Epelde, 2024). The nascent discipline of explainable AI (XAI) seeks to elucidate these opaque systems, yet converting intricate mathematical operations into comprehensible explanations continues to pose significant difficulties (Parker, 2025). Recent evidence indicates that integrating interpretability into model design from the outset, rather than as an afterthought, may be the most advantageous approach (Giacobbe *et al.*, n.d.).

Ethical difficulties are a substantial challenge recognized in the literature. The surveillance tools that augment AI's effectiveness in disease detection also create substantial privacy concerns. These technologies typically collect significant quantities of personal and locational data from sources such as mobile devices and social media platforms (MSc, 2024). Many studies have highlighted ongoing issues over data ownership, adequate consent protocols, and safeguards against exploitation by commercial or governmental entities. Recent polls indicate that public concern about these issues significantly influences the willingness to participate in digital illness monitoring projects (Parker, 2025). Furthermore, accountability mechanisms are insufficiently established for situations in which AI systems err, a scenario that is inevitable in any computational framework. Many authors have noted that a misidentified outbreak signal can trigger unnecessary and costly public health measures or, conversely, fail to recognize a genuine threat with potentially disastrous consequences. The research consistently emphasizes the necessity for thorough ethical frameworks and clear regulatory criteria before the appropriate progression of general use.

Adaptability to novel pathogens emerges in the literature as a critical technical challenge, almost a paradox in infectious disease surveillance. The novelty of emerging pathogens, which makes them dangerous, also makes them difficult targets for AI systems trained on historical data (Ali *et al.*, 2025). Numerous research investigations into the COVID-19 pandemic underscored this issue since predicted models based on prior coronavirus outbreaks faltered in addressing the distinct transmission patterns of novel variants. The literature elucidates that infections perpetually develop, altering their genetic composition, transmission dynamics, and clinical manifestations more rapidly than several AI systems can. This biological arms race requires surveillance systems with comparable adaptability (Ali *et al.*, 2025). Some scholars advocate for continuous monitoring frameworks with real-time data integration; nevertheless, others observe that the implementation of such adaptable systems entails significant technical complexity (MSc, 2024). Recent literature on uncertainty-aware modeling tools indicates intriguing methods for effectively operating with the limited data commonly present during the initial phases of an outbreak (Zhou *et al.*, 2024).

Resource constraints constitute another significant obstacle commonly referenced in the literature. Numerous studies emphasize that the development and maintenance of advanced AI systems necessitate significant financial resources and specialized technical knowledge (MSc, 2024). A thorough evaluation of public health infrastructure in 2023 uncovered significant discrepancies; whereas a select few affluent metropolitan health departments have adopted sophisticated analytics, the majority of rural and underprivileged areas lack even fundamental computational resources (Shi, 2024). Economic assessments reveal that the development of a singular, well-structured AI model generally exceeds $1 million, far surpassing the financial constraints faced by most public health institutions. Numerous recent studies contend that failing to rectify these resource disparities via targeted funding initiatives, technical assistance programs, and collaborative infrastructure development may lead to AI-enhanced surveillance exacerbating, rather than alleviating, existing health disparities among communities (Kenan Institute of Private Enterprise, 2024).

Critical Case Study Analysis of AI-Driven Surveillance Systems

Various organizations and initiatives throughout the United States are actively exploring and implementing AI and machine learning methods for infectious disease surveillance. Examining specific case studies provides valuable insights into the practical application, successes, and challenges related to these approaches.

The Centers for Disease Control and Prevention have recognized the promise of AI and ML via their Data Modernization

Initiative, which seeks to utilize these sophisticated tools for analyzing extensive and intricate public health data (CDC, 2025). These initiatives encompass MedCoder, which employs natural language processing and machine learning to streamline the coding of mortality data, and TowerScout, a web application created in partnership with UC Berkeley to automatically identify cooling towers from satellite imagery, supporting the response to Legionnaires' disease. This tool aids the CDC's response to Legionnaires' disease outbreaks by quickly identifying potential sources of the bacteria (CDC, 2025). The CDC is also exploring the use of AI and ML for various other public health applications, including forecasting trends in opioid overdose mortality, enhancing syndromic surveillance using large language models, and identifying the origins of foodborne illness outbreaks. The CDC acknowledges the significance of health equity and is concentrating on uncovering and addressing possible biases in AI and ML algorithms utilized in their surveillance initiatives. The National Syndromic Surveillance Program (NSSP) at the CDC consolidates electronic health data from more than 6,900 healthcare facilities across 50 states, the District of Columbia, and Guam, typically within 24 hours of a patient visit (CDC, 2024). This initiative is designed to identify, analyze, track, and address public health issues promptly. These initiatives illustrate the CDC's dedication to incorporating AI and ML into their public health surveillance framework to enhance responsiveness and precision.

BlueDot, a Canadian company, has achieved global acclaim for its innovative application of AI in the early identification of infectious disease outbreaks (BlueDot Leverages Data Integration to Predict COVID-19 Spread, n.d.). Significantly, BlueDot was one of the pioneers in detecting the COVID-19 outbreak in Wuhan, China, notifying its clients on December 31, 2019, a full five days before the World Health Organization's formal announcement (BlueDot, n.d.). The platform employs advanced techniques in natural language processing and machine learning to scrutinize extensive global data, such as news articles and airline ticketing information, allowing for precise predictions regarding the dissemination of diseases like Ebola and Zika (BlueDot Leverages Data Integration to Predict COVID-19 Spread, n.d.). The software uses a strict methodology, collecting and evaluating data every 15 minutes to continuously track over 190 different infectious diseases and symptoms in 65 languages. By carefully comparing possible hazards with flight schedules and other data, BlueDot can predict outbreak risks and promptly notify its customers, which include governments, hospitals, and corporations (Amazon Web Services, 2020). BlueDot's innovative solution has reduced manual detection time by almost 90% (Cade, 2024). The correct API endpoint is now visible over 94% of the time when utilizing Cohere Classify (Alder, 2025). The implementation of BlueDot Assistant has led to a significant reduction in weekly horizon scanning activities for surveillance teams, achieving a decrease of up to 88% (Cade, 2024). The "Blue Dot Effect"—the natural tendency to identify dangers even when they are less frequent—is one potential disadvantage of BlueDot, despite its remarkable features (Jones, 2024). The company is constantly working to improve its platform and approach, as evidenced by its efforts to develop its AI models and expand its capabilities. BlueDot is a wonderful example of how artificial intelligence (AI) can be utilized to provide accurate and timely early warnings of the threats that infectious illnesses pose globally.

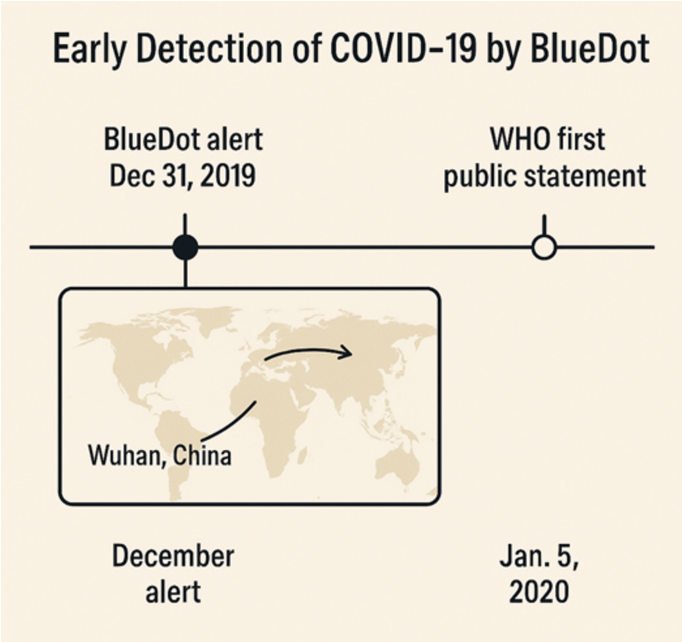

Figure 3. A timeline shows BlueDot's early detection of COVID-19 on December 31, 2019, five days before the World Health Organization issued its first alert. By analyzing global data sources using machine learning and natural language processing, BlueDot was able to predict outbreak hotspots with remarkable speed and accuracy.

HealthMap has been a prominent player in global infectious disease surveillance since 2006; it is a freely accessible, automated electronic information system developed by researchers at Boston Children's Hospital (45). HealthMap aggregates data from various freely available electronic media sources, including online news aggregators like Google News, expert-curated systems like ProMED-mail, and validated official alerts from organizations such as the WHO (Brownstein *et al.*, 2008). The platform employs text-processing algorithms to classify outbreak alerts by location and disease, visualizing this information on an interactive geographic map (Freifeld *et al.*, 2008). HealthMap notably monitored early press and social media reports regarding the Ebola outbreak in West Africa in 2014, which documented a cluster of pneumonia cases in Wuhan in December 2019, before the official acknowledgment of the COVID-19 pandemic (HealthMap, 2024). The assessment of HealthMap's automated classification algorithms indicates accuracy rates ranging from 84% to 91%, highlighting their capability in handling substantial amounts of data (Freifeld *et al.*, 2008). Nonetheless, akin to any automated system, HealthMap encounters certain limitations. The frequency of reports could be shaped more by the perceived economic and social disruption resulting from an outbreak than by its actual morbidity or mortality (Brownstein *et al.*, 2008). Furthermore, the precise classification of alerts that encompass various locations or diseases presents significant challenges (Freifeld *et al.*, 2008). In light of these constraints, HealthMap continues to serve as an important open-source tool for monitoring global infectious diseases in real time.

Table 1. Summary of selected AI/ML-based systems for infectious disease surveillance in the United States

System Name	Methodology	Reported Successes	Identified Limitations	Accessibility
CDC Systems (e.g., MedCoder, TowerScout, BioSense)	NLP, ML, Deep Learning; Analysis of EHRs, satellite imagery, mortality data	Improved speed and accuracy in mortality data coding; Accelerated outbreak response for Legionnaires' disease; Near real-time analysis of syndromic data from over 6,900 facilities	Potential for data biases impacting health equity; Survey data suggests challenges with timeliness and flexibility	Primarily for internal use by public health encies
BlueDot	NLP, ML; Analysis of news, health forums, airline ticketing data	Early warning of COVID-19 outbreak; Tracking of over 190 diseases globally; 88% reduction in manual horizon scanning time reported; 94% accuracy in endpoint identification	Tendency towards identifying problems even when less prevalent ("Blue Dot Effect"); marketing and attendance management issues; technical compatibility requirements	Subscription-based service for governments, businesses, and healthcare organizations
HealthMap	NLP; Aggregation and analysis of news, ProMED-mail, official alerts	Tracking of Ebola and early signals of COVID-19; Global visualization of disease outbreaks; 84-91% accuracy for automated classifier	Frequency of reports may not correlate with severity; Challenges in classifying multi-location outbreaks; Timeliness varies across countries (13-94% for influenza)	Freely accessible web-based platform for public health officials, researchers, and the public

4.3. Recommendations

To improve the effectiveness and impact of artificial intelligence and machine learning in the early detection of emerging infectious diseases in the United States, several important recommendations should be taken into account.

Enhancing the quality and representation of data is essential (Zachariah *et al.*, 2018). It is essential to focus on investing in initiatives that facilitate the gathering of more diverse and representative datasets. This includes data from historically underrepresented populations and low- and middle-income countries, which can serve as vital sources of emerging pathogens (Fahim, 2024). It is crucial to implement strategies that tackle data missingness and inconsistencies across diverse data sources, including standardized reporting protocols and data integration platforms (MSc, 2024). Moreover, advancing data harmonization and standardization will enable the smooth integration of diverse data streams, leading to more thorough and resilient analyses (Ekundayo, 2024).

Enhancing the transparency of AI and ML models is crucial for building trust and promoting their adoption among public health professionals (Epelde, 2024). The continuous exploration and progress in the field of explainable AI (XAI) are crucial for improving the understanding of these complex systems (Giacobbe *et al.*, n.d.). Exploring the combination of inherently interpretable models with advanced deep learning architectures may yield significant benefits. Developing user-friendly tools and interfaces that help public health practitioners understand the factors affecting AI-generated predictions will be essential for their successful application (Ennab & Mcheick, 2024).

Establishing clear ethical guidelines and regulations is essential for the responsible deployment of AI and ML in public health surveillance (Parker, 2025). These frameworks should address critical issues such as data privacy, the necessity of informed consent for data usage, and the assignment of accountability when AI systems are involved in decision-making processes. Prioritizing the development and implementation of privacy-preserving AI techniques, such as federated learning and differential privacy, can help mitigate some of these concerns (Suvvari & Kandi, 2024). Additionally, robust data security measures must be in place to protect sensitive health information from unauthorized access and breaches.

Promoting technical adaptability is essential given the ever-evolving nature of infectious diseases (Zhou *et al.*, 2024). AI and ML models need to be designed for continuous monitoring and improvement to effectively tackle the emergence of new pathogens and variants, as well as changes in disease transmission patterns (HITRUST, 2023). Incorporating real-time data streams and feedback loops into the model training process will allow for ongoing performance improvement. Further research is needed to develop AI and ML techniques that can operate effectively even with the limited data that is typically available during the initial stages of a novel outbreak (Zhou *et al.*, 2024).

Addressing resource constraints is crucial for ensuring that the benefits of AI and ML in early disease detection are widely accessible (AI-Driven Surveillance System to Combat Emerging Infectious Diseases, 2025). Enhanced funding for public health infrastructure is essential to facilitate the development, implementation, and upkeep of these sophisticated technologies. Investigating the implementation of open-source AI platforms and encouraging collaborative efforts among research institutions, government entities, and technology firms can lead to cost reductions and enhance accessibility to these resources (CDC, 2025). Ultimately, it is crucial to allocate resources toward training and capacity-building initiatives that will empower public health professionals with the necessary skills and knowledge to effectively leverage AI and ML technologies in their practices (Hattab *et al.*, 2025).

4.4. Future Research Directions

There are many prospects for more research in the area of using AI and machine learning to detect infectious diseases in the US in a timely manner. Several important areas require concentrated effort to enhance our proficiency in this crucial field.

Different fields must work together more effectively. Experts from a wide range of disciplines, including artificial intelligence, epidemiology, public health, data science, behavioral science, ethics, and healthcare policy, must form strong relationships to make meaningful progress (Zhou *et al.*, 2024). These collaborations will deepen our comprehension of the challenges and support the creation of innovative and practical solutions.

Future research should focus on the integration of multi-scale data (MSc, 2024). Our understanding of disease emergence and transmission could be greatly improved by the creation of AI and ML models that can thoroughly analyze and synthesize data across multiple dimensions, including pathogen genomic characteristics, patient clinical information, environmental factors influencing disease spread, and social determinants of health (MSc, 2024).

The development of robust validation frameworks is crucial (Luna, 2025). Rigorous evaluations are required to ensure the accuracy, dependability, and applicability of AI and ML models for early illness detection in diverse populations and real-world contexts (Luna, 2025). These validation efforts will help to establish confidence in the performance of these technologies and guide their appropriate deployment.

A crucial area of inquiry must center on enhancing the detection of early indicators and emerging pathogens. Methods in artificial intelligence and machine learning are essential for preventing future pandemics by identifying subtle, early indicators of emerging diseases, including those associated with previously unknown or entirely novel pathogens.

Future studies should also assess the impact of AI and ML-driven early detection systems on real public health outcomes (SCIP, 2024). Studies are required to evaluate the effects of these technologies on decreasing morbidity, mortality, and the general transmission of infectious diseases (SCIP, 2024).

Ultimately, working together on a global scale is crucial. Given the global nature of emerging infectious disease threats, fostering collaboration among experts and public health organizations worldwide is essential for sharing data, knowledge, and best practices in the application of AI and ML, which will be critical for strengthening global health security.

5. CONCLUSION

The early detection of newly developing infectious diseases in the US could be revolutionized by artificial intelligence and machine learning. These technologies provide effective tools to improve public health surveillance and response capabilities by enabling the quick analysis of large, diverse datasets, greatly increasing the accuracy of predictive models (e.g., reaching over 90% accuracy in some applications), and enabling real-time monitoring of possible outbreaks. The successful and responsible application of AI and ML in this vital field requires a comprehensive understanding and careful management of inherent obstacles and constraints. We must meticulously address concerns regarding data biases, the interpretability of intricate models, ethical implications of data privacy and utilization, the technical adaptability of systems to emerging dangers, and the resource investments necessary for development and implementation.

To effectively harness the advantages of AI and ML in protecting the nation from developing infectious diseases, it is essential to concentrate on three critical domains. These include making the data used to train AI models more representative and high-quality, making the models more transparent and easy for public health professionals to understand, creating strong regulatory frameworks and clear ethical guidelines to control their use, encouraging the technical adaptability of AI systems to new and changing threats, and resolving resource limitations that might prevent their widespread adoption. By thoughtfully and strategically leveraging the power of AI and machine learning, the United States can significantly strengthen its defenses against the ongoing threat of emerging infectious diseases and better protect the health and well-being of its population.

REFERECNCES

Accuray. (n.d.). *Overcoming AI Bias: Understanding, Identifying and Mitigating Algorithmic Bias in Healthcare—Accuray.* Retrieved March 27, 2025, from https://www.accuray.com/blog/overcoming-ai-bias-understanding-identifying-and-mitigating-algorithmic-bias-in-healthcare/

Jones, A. (2024, October). *AI Alignment June 2024 course retrospective – BlueDot Impact.* https://bluedot.org/blog/ai-alignment-june-2024-retro/

Ahmad, F. B. (2024). *Mortality in the United States—Provisional Data, 2023.* MMWR. Morbidity and Mortality Weekly Report, 73. https://doi.org/10.15585/mmwr.mm7331a1

AI-driven surveillance system to combat emerging infectious diseases. (2025, January 2). News-Medical. https://www.news-medical.net/news/20250102/AI-driven-surveillance-system-to-combat-emerging-infectious-diseases.aspx

Aiforia. (2021, October 21). *Case study: Infectious diseases and AI.* https://www.aiforia.com/resource-library/infectious-diseases-and-ai

Alder, S. (2025, March 19). *The Biggest Healthcare Data Breaches of 2024.* The HIPAA Journal. https://www.hipaajournal.com/biggest-healthcare-data-breaches-2024/

Ali, H. H., Ali, H. M., Ali, H. M., Ali, M. A., Zaky, A. F., Touk, A. A., Darwiche, A. H., & Touk, A. A. (2025). *The Role and Limitations of Artificial Intelligence in Combating Infectious Disease Outbreaks.* Cureus. https://doi.org/10.7759/cureus.77070

Amazon Web Services. (2020). *BlueDot Case Study.* Amazon Web Services, Inc. https://aws.amazon.com/solutions/case-studies/bluedot

BlueDot leverages data integration to predict COVID-19 spread. (n.d.). FME by Safe Software. Retrieved March 27, 2025, from https://fme.safe.com/fme-in-action/customers/bluedot

BlueDot. (n.d.). *Customer case studies & testimonials.* BlueDot. Retrieved March 27, 2025, from https://bluedot.global/customer-stories

Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandl, K. D. (2008). Surveillance Sans Frontières: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project. *PLoS Medicine, 5*(7), e151. https://doi.org/10.1371/journal.pmed.0050151

Cade, J. (2024, August 7). *BlueDot unveils global infectious disease surveillance solution.* BlueDot. https://bluedot.global/bluedot-unveils-next-gen-global-infectious-disease-surveillance-solution-cutting-manual-detection-time-by-nearly-90

CDC Foundation. (2024, August). *The Costs of Getting Sick | Contagious Conversations.* CDC Foundation. https://www.cdcfoundation.org/conversations/cost-of-getting-sick

CDC. (1994, April). *Addressing Emerging Infectious Disease Threats: A Prevention Strategy for the United States Executive Summary.* https://www.cdc.gov/mmWR/preview/mmwrhtml/00031393.htm

CDC. (2024, July). *The national syndromic surveillance program.* https://www.cdc.gov/nssp/documents/NSSP-Infographic.pdf

CDC. (2025, February 21). *Artificial Intelligence and Machine Learning, Technologies.* CDC. https://www.cdc.gov/surveillance/data-modernization/technologies/ai-ml.html

Center for Global Digital Health Innovation. (n.d.). *Predictive Data Analytics in Global Health.* Center for Global Digital Health Innovation. Retrieved March 27, 2025, from https://publichealth.jhu.edu/center-for-global-digital-health-innovation/september-2023-predictive-data-analytics-in-global-health

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

CIDRAP. (2024, February 16). *US Lyme Disease infections.* https://www.cidrap.umn.edu/lyme-disease/documented-us-lyme-disease-infections-soared-2022-after-updated-case-definition

Dankwa-Mullan, I. (2024). Health Equity and Ethical Considerations in Using Artificial Intelligence in Public Health and Medicine. *Preventing Chronic Disease, 21.* https://doi.org/10.5888/pcd21.240245

Ekundayo, F. (2024). Using machine learning to predict disease outbreaks and enhance public health surveillance. *World Journal of Advanced Research and Reviews, 24*(3), 794–811. https://doi.org/10.30574/wjarr.2024.24.3.3732

Emerging infectious disease. (2025). In *Wikipedia.* https://en.wikipedia.org/w/index.php?title=Emerging_infectious_disease&oldid=1280915963

Ennab, M., & Mcheick, H. (2024). Enhancing interpretability and accuracy of AI models in healthcare: A comprehensive review on challenges and future directions. *Frontiers in Robotics and AI, 11*, 1444763. https://doi.org/10.3389/frobt.2024.1444763

Epelde, F. (2024). How AI Could Help Us in the Epidemiology and Diagnosis of Acute Respiratory Infections? *Pathogens, 13*(11), Article 11. https://doi.org/10.3390/pathogens13110940

Fahim, S. (2024, July 23). *Artificial Intelligence in Disease Detection and Prevention.* Valparaiso University. https://onlinedegrees.valpo.edu/ai-in-disease-prevention/

Freifeld, C. C., Mandl, K. D., Reis, B. Y., & Brownstein, J. S. (2008). HealthMap: Global Infectious Disease Monitoring through Automated Classification and Visualization of Internet Media Reports. *Journal of the American Medical Informatics Association, 15*(2), 150–157. https://doi.org/10.1197/jamia.M2544

Ghaffar Nia, N., Kaplanoglu, E., & Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discover Artificial Intelligence, 3*(1), 5. https://doi.org/10.1007/s44163-023-00049-5

Giacobbe, D. R., Zhang, Y., & de la Fuente, J. (n.d.). Explainable artificial intelligence and machine learning: Novel approaches to face infectious diseases challenges. *Annals of Medicine, 55*(2), 2286336. https://doi.org/10.1080/07853890.2023.2286336

Hattab, G., Irrgang, C., Körber, N., Kühnert, D., & Ladewig, K. (2025). The Way Forward to Embrace Artificial Intelligence in Public Health. *American Journal of Public Health, 115*(2), 123-128. https://doi.org/10.2105/AJPH.2024.307888

HealthMap. (2024). In *Wikipedia.* https://en.wikipedia.org/w/index.php?title=HealthMap&oldid=1235220506

HITRUST. (2023, November 15). *The Ethics of AI in Healthcare.* https://hitrustalliance.net/blog/the-ethics-of-ai-in-healthcare

Jeong, J., Kim, S., Pan, L., Hwang, D., Kim, D., Choi, J., Kwon, Y., Yi, P., Jeong, J., & Yoo, S.-J. (2025). Reducing the workload of medical diagnosis through artificial intelligence: A narrative review. *Medicine, 104*(6), e41470. https://doi.org/10.1097/MD.0000000000041470

Luna, M. (2025, March). *Machine Learning in Epidemiology: Enhancing Disease Forecasting and Outbreak Prevention through AI.*

Kenan Institute of Private Enterprise. (2024, January 12). *Is AI the Answer to Lowering Healthcare Costs?* - Frank Hawkins Kenan Institute of Private Enterprise. https://kenaninstitute.unc.edu/news-media/is-ai-the-answer-to-lowering-healthcare-costs/

MacIntyre, C. R., Chen, X., Kunasekaran, M., Quigley, A., Lim, S., Stone, H., Paik, H., Yao, L., Heslop, D., Wei, W., Sarmiento, I., & Gurdasani, D. (2023). Artificial intelligence in public health: The potential of epidemic early warning systems. *The Journal of International Medical Research, 51*(3), 03000605231159335. https://doi.org/10.1177/03000605231159335

Malesu, V. K. (2025, March 3). Can AI Outperform Doctors in Diagnosing Infectious Diseases? *News-Medical.* https://www.news-medical.net/health/can-ai-outperform-doctors-in-diagnosing-infectious-diseases.aspx

McArthur, D. B. (2019). Emerging Infectious Diseases. *The Nursing Clinics of North America, 54*(2), 297–311. https://doi.org/10.1016/j.cnur.2019.02.006

McCormick, B. (2025, January 19). *AI Outperforms Traditional Methods in Predicting Ovarian Cancer Surgery Outcomes.* AJMC. https://www.ajmc.com/view/ai-outperforms-traditional-methods-in-predicting-ovarian-cancer-surgery-outcomes

Melchane, S., Elmir, Y., Kacimi, T., & Boubchir, L. (2024, November 14). *Artificial Intelligence for Infectious Disease Prediction and Prevention: A Comprehensive Review.* https://arxiv.org/html/2411.10486v1

MSc, A. D. B. (2024, March 18). The Role of AI in Pathogen Detection and Epidemic Prediction. *News-Medical.* https://www.news-medical.net/health/The-Role-of-AI-in-Pathogen-Detection-and-Epidemic-Prediction.aspx

Olaboye, J. A., Maha, C. C., Kolawole, T. O., & Abdul, S. (2024). Innovations in real-time infectious disease surveillance using AI and mobile data. *International Medical Science Research Journal, 4*(6), 647-667. https://doi.org/10.51594/imsrj.v4i6.1190

Omar, M., Brin, D., Glicksberg, B., & Klang, E. (2024). *Utilizing Natural Language Processing and Large Language Models in the Diagnosis and Prediction of Infectious Diseases: A Systematic Review* (p. 2024.01.14.24301289). medRxiv. https://doi.org/10.1101/2024.01.14.24301289

Parums, D. V. (2023). Editorial: Infectious Disease Surveillance Using Artificial Intelligence (AI) and its Role in Epidemic and Pandemic Preparedness. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 29*, e941209-1-e941209-4. https://doi.org/10.12659/MSM.941209

Parker, M. (2025, March). *Ethical Challenges in the use of AI for Infectious Disease Epidemiology: A Double-Edged Sword.* https://www.ethox.ox.ac.uk/blog/ethical-challenges-in-the-use-of-ai-for-infectious-disease-epidemiology-a-double-edged-sword

SCIP. (2024, October 28). *Predictive Analytics in Healthcare: The Future of Disease Prevention—Strategic Consortium of Intelligence Professionals (SCIP).* https://www.scip.org/news/685370/Predictive-Analytics-in-Healthcare-The-Future-of-Disease-Prevention-.htm

Shi, E. K. (2024, October). *AI in Public Health: Gaps, Disparities, and Remarkable Potential.* Global Health NOW. https://globalhealthnow.org/2024-10/ai-public-health-gaps-disparities-and-remarkable-potential

Srivastava, V., Kumar, R., Wani, M. Y., Robinson, K., & Ahmad, A. (2025). Role of artificial intelligence in early diagnosis and treatment of infectious diseases. *Infectious Diseases, 57*(1), 1–26. https://doi.org/10.1080/23744235.2024.2425712

Suvvari, T. K., & Kandi, V. (2024). Artificial intelligence enhanced infectious disease surveillance—A call for global collaboration. *New Microbes and New Infections, 62*, 101494. https://doi.org/10.1016/j.nmni.2024.101494

Westat. (n.d.). *Can natural language processing improve the completeness of immunization data?* Westat. Retrieved March 26, 2025, from https://www.westat.com/projects/natural-language-processing-improve-immunization-data/

Zachariah, P., Hill-Ricciuti, A., & Natarajan, K. (2018). 1765. Use of a Natural Language Processing-Based Informatics Pipeline for Infectious Disease Syndrome Surveillance. *Open Forum Infectious Diseases, 5*(Suppl 1), S63–S64. https://doi.org/10.1093/ofid/ofy209.150

Zhou, H.-Y., Li, Y., Li, J.-Y., Meng, J., & Wu, A. (2024). Harnessing the power of artificial intelligence to combat infectious diseases: Progress, challenges, and future outlook. *The Innovation Medicine, 2*(4), 100091–15. https://doi.org/10.59717/j.xinn-med.2024.100091