Outbreak, Surveillance, Investigation & Response (OSIR) Journal
Field Epidemiology Training Program, Division of Epidemiology
Department of Disease Control, Ministry of Public Health, Thailand
Tel: +6625903894, Fax: +6625903845, Email: osireditor@osirjournal.net, https://he02.tci-thaijo.org/index.php/OSIR

Enhancing Digital Disease Surveillance in Thailand Using Information Technology, Data Engineering, Data Science, and Artificial Intelligence

Suphanat Wongsanuphat*, Jaruwan Malaikham, Supansa Suriya, Jiraporn Prommongkhol, Narawadee Khampha, Thapanee Khempetch

Division of Epidemiology, Department of Disease Control, Ministry of Public Health, Thailand

*Corresponding author, email address: suphanat.wong@gmail.com

Received: 4 Nov 2024; Revised: 1 Dec 2024; Accepted: 13 Jan 2025
https://doi.org/10.59096/osir.v18i1.272056

Abstract

Recent advancements in data engineering, data science, and artificial intelligence have revolutionized disease surveillance systems globally. This study examines the implementation of these advancements in Thailand. We integrated these advancements to enhance the four key steps of public health surveillance: data collection, data analysis, data interpretation, and data dissemination. We expanded data collection to include data environment and integration, designing systems to manage multiple sources and facilitating seamless integration. To support analysis and interpretation, we adopted a design thinking approach and developed intuitive tools for exploring disease situations. We identified target users and described the data distribution mechanism. We integrated three major databases: digital disease surveillance, syndromic surveillance, and event-based surveillance, all managed by the Department of Disease Control. Data environments were divided into clusters for extraction, integration, and a data mart for specific use cases. Automated hourly extract-transform-load processes using Apache Airflow facilitated real-time data integration, ensuring seamless data management and timely updates. Data analysis solutions, including automated validation algorithms and business intelligence tools with user-friendly interfaces, were developed according to findings from the design thinking workshop. We developed open data published on dashboards and closed data managed through the Digital Export System. Artificial intelligence-enhanced early warning systems provided notifications of an outbreak to public health authorities via the instant messaging application LINE. In conclusion, the integration of information technology, data engineering, and data science has significantly enhanced Thailand's disease surveillance system, improving data collection, analysis, interpretation, and dissemination of results, leading to more efficient public health responses.

Keywords: disease surveillance, information technology, data engineering, data science, artificial intelligence

Introduction

Public health surveillance is defined as “the ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice”.[1] In recent years, the convergence of information technology, data engineering, and data science has revolutionized various sectors, including healthcare and public health.

Specifically, the application of these advancements has significantly enhanced disease surveillance systems, enabling quicker detection, more accurate prediction, and more effective responses to public health threats.[2] Real-time data collection from hospitals, clinics, laboratories, and wearable devices ensures immediate availability of health data, facilitating faster identification of potential outbreaks.[3] Automated reporting reduces delays associated with manual data entry, keeping information up-to-date and readily accessible.[4] With the advent of advanced analytical techniques, including machine learning and statistical modeling, vast amounts of health data can now be analyzed to help researchers identify patterns and trends, predicting disease outbreaks by considering factors such as population density, climate conditions, travel patterns, and historical data.[3,5] This integration of technology, data science, and artificial intelligence has transformed traditional surveillance methods, allowing for early detection of emerging threats and enabling data-driven decision-making for efficient public health responses.[6–9]

In the context of Thailand, a country that faces diverse health challenges ranging from infectious diseases to chronic conditions, leveraging digital tools for disease surveillance holds immense potential for improving public health outcomes. Thailand, known for its robust healthcare infrastructure and proactive approach to public health, has recognized the importance of integrating information technology and data science into its disease surveillance system. By harnessing the power of digital platforms, real-time data collection, and advanced analytics, Thailand aims to strengthen its ability to monitor, analyze, and respond to disease outbreaks promptly and efficiently.

Disease surveillance in Thailand is primarily facilitated through the R506 reporting system, in accordance with the Communicable Diseases Act B.E. 2558 (2015), using a comprehensive framework for collecting and analyzing notifiable diseases under the surveillance system across the country.[10] This system enables healthcare providers to report cases of notifiable diseases, according to their case definition, facilitating timely responses to potential outbreaks.[11] The R506 reporting system is complemented by digital tools that enhance data accuracy and accessibility, further supporting public health initiatives. A notable example of Thailand's efforts to enhance its disease surveillance system was the implementation of information technology during the coronavirus disease 2019 (COVID-19) pandemic.[12,13]

This study aims to explore the landscape of digital disease surveillance (DDS) in Thailand, highlighting the key strategies, technologies, and initiatives employed to enhance public health monitoring and response capabilities. It delves into the role of information technology and data science in transforming traditional surveillance methods, facilitating the early detection of emerging threats, and enabling data-driven decision-making at local, regional, and national levels.

Methods

Public health surveillance consists of four crucial steps: data collection, data analysis, data interpretation, and data dissemination.[1] In this study, we integrated information technology, data engineering, and data science to enhance these steps in the surveillance system. The research questions and methods to enhance digital disease surveillance using information technology, data engineering, and data science for each step are described in Table 1. We also describe recent implementations, challenges, and suggest opportunities for improvement.

Table 1. Summary of surveillance steps, research questions, and methods to enhance digital disease surveillance in Thailand using information technology, data engineering, and data science

Surveillance step	Research question	Method
Data collection, data environment, and data integration	What are the data sources that the user needs for disease surveillance?	Describing the current surveillance system.
	How to design and implement a multiple data source integration system?	Designing a data environment and automating data management pipeline.
Data analysis and interpretation	What are the needs of users for disease surveillance data analysis?	Developing a user-friendly surveillance data analysis system.
Data dissemination	How to design an effective and timely data dissemination system?	Identifying target data users. Designing data dissemination pathways.

Definitions

Information Technology

The use of computers, networks, and software to manage and process information, enabling data storage, retrieval, and communication.[14]

Data Engineering

The practice of designing and building systems for collecting, storing, and analyzing data at scale, including creating data pipelines and infrastructure.[15]

Data Science

Data science combines statistics, mathematics, and computer science to extract meaningful insights from data using algorithms, machine learning, and visualization.[16]

Artificial Intelligence (AI)

Artificial intelligence is a system capable of performing tasks that typically require human intelligence, such as learning, decision-making, and language understanding.[17]

Data Collection, Data Environment, and Data Integration

We identified diverse data sources related to disease surveillance by describing current surveillance systems in the Thai Department of Disease Control. We designed data environments to manage data from multiple sources, facilitating seamless integration and aggregation into a centralized surveillance system. We integrated these diverse data sources using real-time data pipeline automation (Figure 1).

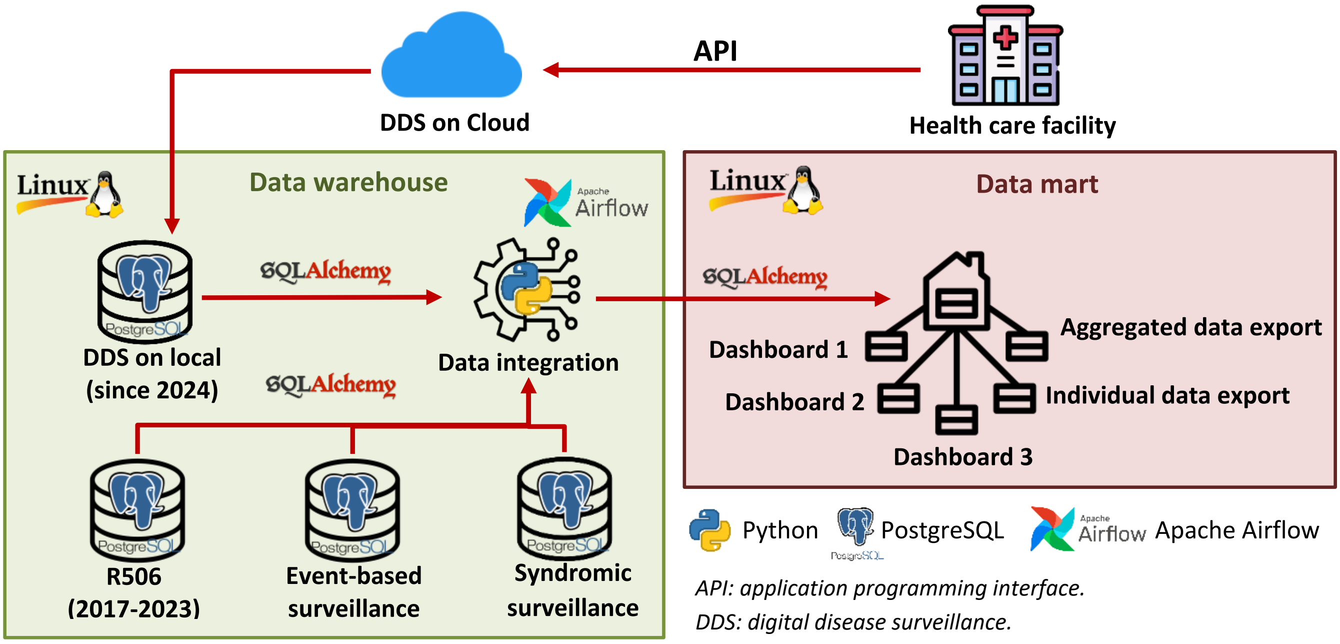

Figure 1. Data collection, data environment, and data integration pipeline of the digital disease surveillance system

Data Analysis and Interpretation

In our previous study, we applied a design thinking approach to gather data analysis requirements for 50 epidemiological staff, 10 executives, and 10 frontline staff at the national office, regional health office, provincial health office, hospital, and sub-district hospital levels in Chiang Mai, Uthai-Thani, and Nakhon Sawan provinces.[18] In this step, our goal was to create user-friendly tools to support the analysis and interpretation of surveillance data based on user requirements mentioned in our previous study.[18] We identified the potential of information technology, data engineering, and data science techniques based on our previous study.[18] These tools are designed to enable public health officials, executives, and frontline staff to explore data trends, disease clusters, and hotspots intuitively. Additionally, we designed tools for user-friendly data interpretation to generate hypotheses based on the findings of the data analysis.

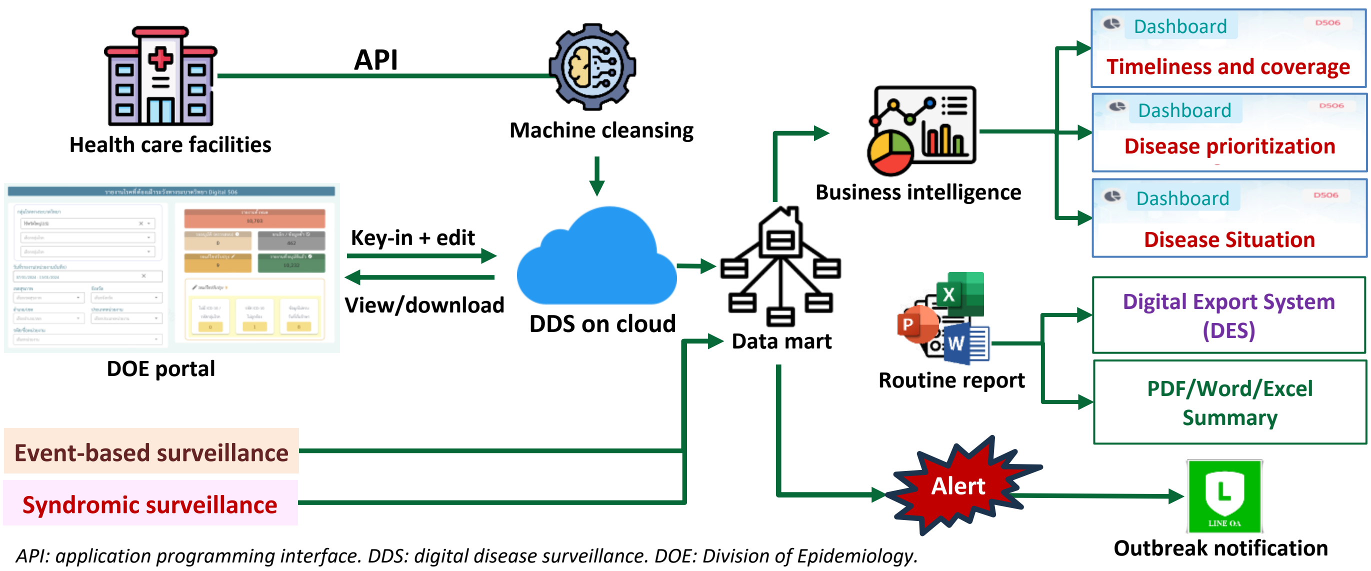

Figure 2. Data analysis, interpretation, and dissemination pipeline of digital disease surveillance

Data Dissemination and Response

Potential data users under data dissemination were identified and classified. Data dissemination pathways were described. We also established early warning systems by identifying appropriate algorithms that generate early warning signals based on predictive analytics and artificial intelligence. We evaluated reporting performance using data from the DDS system (2024) and the R506 system (2023). Reporting coverage was defined as the proportion of hospitals submitting data and timeliness as the percentage of records reported within seven days of diagnosis. Data from 1,409 eligible hospitals were analyzed using consistent criteria for both systems to compare reporting coverage and timeliness.

Ethics

Ethics approval was not required for this work as it is part of the routine mission of the Division of Epidemiology. This study did not collect individual data nor involve any personal sensitive data.

Results

In Thailand, data engineering, data science, and artificial intelligence were integrated into data collection platforms, data integration, data analysis, data interpretation, and data dissemination of public health surveillance.

Data Collection, Data Environment, and Data Integration

We reviewed three current disease surveillance systems: a case-based surveillance system, a syndromic surveillance system, and an event-based surveillance system. The case-based disease surveillance system collects individual data from hospitals across the country. Data from this system is divided into two phases: 1) R506 (before 2024) and 2) the DDS system (2024 and after). This surveillance system is a major method in Thailand, which seamlessly connects individual data from 1,409 healthcare facilities, including public and private hospitals. R506 contains weekly offline disease surveillance data sent by provincial health offices, while the DDS contains real-time disease surveillance data sent directly from hospitals via an application programming interface (API).[19] Fifty-seven notifiable diseases were reported to the DDS consisting of 2,159,925 records as of December 2024. The syndromic surveillance system contains seven groups of symptoms, including influenza-like illnesses, fever, viral conjunctivitis, viral exanthem, acute diarrhea, acute flaccid paralysis, and adverse events following immunization. Data from this surveillance system was aggregated daily at the hospital level. Data from all three surveillance systems are managed by the Thai Department of Disease Control, Ministry of Public Health. Additionally, data from the Thai Health Data Center, which includes demographic data, underlying clinical data, treatment data, and immunization data, were incorporated. There were opportunities to enhance the surveillance systems by including various data sources such as animal health data, vector data, environmental data, and social media feeds. Here, we included three major data sources, namely the DDS, the syndromic surveillance system, and the event-based surveillance system.

For data environment, initially, we set up local environments (on premises) using the Linux operating system. The local environments were divided into three clusters: 1) data warehouse for data extraction and ingestion, 2) data warehouse for data source integration, and 3) data mart for specific use cases such as creating a dashboard. All three local environments were seamlessly interconnected under the Linux operating system. A PostgreSQL database was installed on these local environments.

For data integration, we integrated three data sources using automated hourly batch procedures. Data engineering techniques, specifically extract, transform, and load (ETL), were implemented. First, DDS on a cloud using NoSQL were extracted and transformed to SQL structures using the pymongo and pandas packages in Python. Data were loaded into the DDS on the local environment. Second, data from R506 and the DDS system were extracted using the SQLAlchemy package and then over 10 million rows of surveillance data spanning six years were manipulated and transformed using Python.[37] In this step, data were transformed into multiple data structures based on their specific purposes. Some of the data were integrated with syndromic and event-based surveillance data. Finally, data were loaded back into the data mart using SQLAlchemy. The ETL process was scheduled as an hourly batch using Apache Airflow. A summary of all of these steps is shown in Table 2.

Data Analysis and Interpretation

We designed a wireframe of business intelligence for disease surveillance data analysis based on the aforementioned data analysis requirements using Microsoft PowerPoint, which entails a human-centric approach and data visualization best practices. We prototyped user-friendly interfaces of business intelligence using Tableau.[38] Data were automatically transferred to the Tableau Server using the Tableau Server REST API and Apache Airflow (Table 2). To validate the performance of our tools, we tested our prototypes with the same stakeholders as described in our previous study.[18]

Data Dissemination and Response

We separated data dissemination into two types, namely open data and closed data. To facilitate open data dissemination, user-friendly business intelligence interfaces were developed and integrated into the dashboard section of https://ddsdoe.ddc.moph.go.th/ddss/, as illustrated in Figure 2. Our open data dissemination consisted of 1) timeliness and coverage monitoring business intelligence, 2) disease prioritization business intelligence, and 3) disease situation business intelligence. For closed data, we developed a Digital Export System (DES). First, we indicated data accessibility among different users. Second, we designed and developed a data mart for data dissemination. The data were anonymized to protect personal data privacy. Data anonymization procedures included data deletion, data encryption, and data aggregation. For data encryption and deletion, we deleted or encrypted identifiers such as first name, surname, and citizen identification number. The choice between deletion or encryption depended on the data accessibility level of users. Additionally, we established an "outbreak notification" feature by identifying appropriate algorithms to generate early warning signals with artificial intelligence-enhanced time-series forecasting using the autogluon.timeseries package in Python. This includes developing an outbreak early warning system that automatically flags unusual deviations from predicted trends in surveillance data, e.g., higher than 80, 90, or 95% of the upper prediction interval, with alert mechanisms implemented to notify public health authorities of potential outbreaks or emerging threats in real-time via LINE (Table 2). As of December 2024, the reporting coverage of the DDS system stands at 1,356/1,409 hospitals (96.24%), compared to the coverage of the R506 system in 2023, which had a coverage of 1,239/1,409 hospitals (88.12%). Additionally, the timeliness of reporting to the surveillance system within seven days of the diagnosis date for the DDS system was 83.91%, compared to the performance of R506 in 2023 (74.97%).

Table 2. Summary of the use case of information technology, data engineering, data science, and artificial intelligence in the surveillance steps of data collection, analysis, and dissemination

Surveillance steps	Technology	How the technology is used
Data collection, data environment, and data integration	Information technology	APIs were used to collect real-time data (DDS) and set up local Linux-based environments with PostgreSQL for storage and integration.
	Data engineering	ETL processes such as Python, SQLAlchemy, and Apache Airflow were used for automated data extraction, transformation, and integration of DDS, syndromic, and event-based surveillance data.
	Data science	Data manipulation and transformation using Python packages (e.g., pandas, numpy, pyspark) to harmonize and structure large datasets for analysis.
Data analysis and interpretation	Data engineering	Automated data transfers to Tableau Server via REST APIs and Airflow for visualization prototyping and stakeholder feedback.
	Data science	Prototyping and designing user-friendly data visualization interfaces for disease surveillance using Tableau, focusing on actionable insights.
	Artificial intelligence	Developing early warning systems using AI-based time-series forecasting (e.g., autogluon.timeseries) to predict and flag unusual surveillance trends.
Data dissemination and response	Information technology	Developed open and closed data dissemination systems via dashboards and secure Digital Export Systems with data privacy protocols (e.g., PDPA-compliant anonymization).
	Artificial intelligence	Integrated AI to detect outbreak signals by analyzing trends and deviations, with alert mechanisms through platforms like LINE for real-time public health response.

AI: artificial intelligence. API: application programming interface. DDS: digital disease surveillance. ETL: extract, transform, and load. PDPA: Personal Data Protection Act.

Discussion

Our study established a data integration pipeline of three major disease surveillance systems in Thailand, integrating data from 1,409 healthcare facilities. Together, these three systems facilitated a comprehensive and centralized surveillance system, enhancing real-time data management and integration. Similar initiatives have been implemented globally, such as the BioSense platform in the United States, which also integrates multiple data sources for public health surveillance.[20] However, our study included diverse data types, such as syndromic and event-based data, which provided a broader context for disease monitoring. This comprehensive approach allows for more holistic surveillance but also presents challenges in data harmonization and privacy protection, underscoring the importance of a robust data governance framework.

Combining data engineering and data science has been a significant step forward in real-time disease surveillance systems by enabling automated data pipelines. Data engineering involves multiple steps, including data collection, establishing a data environment, and data integration. Surveillance data from the data warehouse were extracted, transformed, and loaded into a data mart automatically. Data engineering techniques have also been applied to other public health fields, including cancer clinical genomic analysis and women's imaging.[21,22] With a data science approach, large amounts of data can be managed quickly and efficiently. This approach is important when dealing with large-scale epidemics such as COVID-19.[23]

The design thinking approach that we adopted involved epidemiological staff and executives gathering user requirements, resulting in the development of automated validation algorithms and user-friendly interfaces. This human-centric method ensured that the tools met the actual needs of the end-users, improving the efficiency and accuracy of data analysis.[24,25] Studies also emphasized the importance of a user-centric design in public health informatics, highlighting the fact that stakeholder engagement leads to better tool adoption and satisfaction.[26,27] In addition, a previous literature review on developing data visualization for hospital-based surveillance highlighted key issues for tool development, including 1) clear objectives of tools and use case, 2) report content, and 3) interactive use on the screen.[28] Our results align with these findings, demonstrating that gathering comprehensive requirements from users, iterative development, and testing with end-users are critical for successful implementation. The engagement process facilitated the identification of practical challenges and the development of tailored solutions, reinforcing the value of design thinking in public health technology projects.

Our AI approach was embedded in an early warning system resulting in timely alerts for potential outbreaks. Various studies on AI-enhanced early warning systems illustrated promise for their development in the context of disease outbreaks.[29–32] While advanced models such as deep learning provide high accuracy, they pose challenges in interpretability and resource demands. However, AI-based early warning systems must address challenges related to data quality, model explainability, bias mitigation, adaptability, and continuous monitoring, as well as data volume, velocity, variety, availability, and granularity.[33]

Our integration of real-time alert mechanisms via the LINE application is relatively accessible and addresses the need for rapid communication in public health emergencies as implemented in other countries such as Australia.[34,35] Consistent with a study from Canada, a COVID-19 alert application showed high ratios of averted cases and deaths.[36] The successful implementation of these systems in Thailand demonstrates that tailored and context-specific solutions can significantly enhance disease surveillance and response, aligning with global best practices while addressing local needs and challenges.

Limitations

This surveillance system development has some limitations. First, we primarily focused on integrating data from existing surveillance systems, which may not account for gaps in community-level data or non-traditional sources such as social media and environmental data, limiting the system's comprehensiveness. Second, the user requirements for tool design were gathered from a relatively small sample of stakeholders in specific provinces, which may reduce the generalizability of the findings to other regions or levels of the healthcare system. Lastly, while we incorporated predictive algorithms, we did not include a robust validation framework to assess their performance under varying conditions, potentially limiting their reliability in diverse scenarios.

Public Health Recommendations

To optimize real-time disease surveillance, organizations seeking to enhance their systems, similar to the Department of Disease Control, can adopt a structured, step-by-step approach. First, investing in robust IT infrastructure is critical, including scalable cloud-based databases such as PostgreSQL or NoSQL for data storage and processing, and APIs to ensure seamless real-time data exchange between hospitals and central surveillance systems. Next, targeted capacity-building programs should be implemented to train personnel in learning essential skills, such as data engineering techniques (e.g., ETL processes for data integration), data science methods for statistical analysis, visualization using tools such as Tableau and Python, and AI techniques for deploying predictive algorithms to improve outbreak detection. Automated ETL pipelines should then be developed to integrate syndromic, event-based, and case-based surveillance data, with tools such as Apache Airflow enabling frequent updates to maintain data timeliness and accuracy. To enhance analytical capabilities, organizations should design user-friendly dashboards using platforms such as Tableau, incorporating AI-driven early warning systems to flag anomalies and provide predictive insights on disease trends. Real-time communication mechanisms, such as instant messaging applications or platforms such as LINE, should be integrated to immediately notify public health authorities of potential outbreaks. Additionally, organizations must prioritize data privacy by implementing secure dissemination systems that anonymize sensitive information in compliance with existing regulations.

Future studies should integrate additional data sources, such as animal health, vector, environmental, and social media data, to enhance predictive models for emerging health threats. Comparative research on AI algorithms can identify the most effective tools for real-time disease forecasting. In addition, surveillance evaluation should be conducted to monitor performance of enhanced surveillance system.

Conclusion

We have established an integrated data pipeline for disease surveillance in Thailand, combining data from three surveillance systems. This centralized approach enhances real-time data management, providing a comprehensive view of public health monitoring. By employing a design thinking approach, we ensured the system met user needs, improving data accuracy and operational efficiency. The inclusion of AI-driven early warning systems and real-time alerts strengthened outbreak response, aligning with global best practices while addressing local challenges. Our findings underscore the value of robust data governance, user-centered design, and context-specific solutions to enhance public health surveillance and response.

Acknowledgements

We would like to thank the Epidemiological Information Unit and the Intelligence and Non-communicable Diseases Surveillance Development Unit at the Division of Epidemiology, Division of Digital Disease Control, Department of Disease Control, for their technical support. Special thanks to Dr. Sopon Iamsirithaworn, Dr. Chakrarat Pittayawonganon, Dr. Wichan Bhunyakitikorn, Dr. Darin Areechokchai, Dr. Panithee Thammawijaya, and Dr. Seesai Yeesoonseang for providing valuable advice.

Funding

This study was conducted without any financial support or funding.

Conflict of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

Suggested Citation

Wongsanuphat S, Malaikham J, Suriya S, Prommongkhol J, Khampha N, Khempetch T. Enhancing digital disease surveillance in Thailand using information technology, data engineering, data science, and artificial intelligence. OSIR. 2025 Mar;18(1):52–60. doi:10.59096/osir.v18i1.272056.

References

1. Centers for Disease Control and Prevention (US). Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics [Internet]. 3rd edition. [cited 2024 Nov 4]. <https://archive.cdc.gov/#/details?url=https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section1.html>

2. Wang L, Liu Y, Chen H, Qiu S, Liu Y, Yang M, et al. Search-engine-based surveillance using artificial intelligence for early detection of coronavirus disease outbreak. J Big Data. 2023 Nov 11;10:169. doi:10.1186/s40537-023-00847-9.

3. Obe OO, Sarumi OA, Adebayo A. Enhancing epidemiological surveillance systems using dynamic modeling: a scoping review. Proceedings of the 13th International Conference on Soft Computing and Pattern Recognition; 2021 Dec 15–17; Seattle (WA), United States. Cham (CH): Springer; 2022. p. 512–23. doi:10.1007/978-3-030-96302-6_48.

4. Anjaria P, Asediya V, Bhavsar P, Pathak A, Desai D, Patil V. Artificial intelligence in public health: revolutionizing epidemiological surveillance for pandemic preparedness and equitable vaccine access. Vaccines. 2023 Jun 26;11(7):1154. doi:10.3390/vaccines11071154.

5. European Centre for Disease Prevention and Control. Digital technologies for the surveillance, prevention and control of infectious diseases - a scoping review of the research literature [Internet]. Stockholm: European Centre for Disease Prevention and Control; 2021 [cited 2024 Jul 30]. <https://www.ecdc.europa.eu/en/publications-data/digital-technologies-surveillance-preven tion-and-control-infectious-diseases>

6. Fisher S, Rosella LC. Priorities for successful use of artificial intelligence by public health organizations: a literature review. BMC Public Health. 2022;22(1):2146. doi:10.1186/s12889-022-14422-z.

7. Centers for Disease Control and Prevention (US). Artificial Intelligence and Machine Learning: Technologies. [Internet]. Atlanta: Centers for Disease Control and Prevention; 2023 [cited 2025 Jan 15]. <https://www.cdc.gov/surveillance/technologies/index.html>

8. Zhang L, Guo W, Lv C. Modern technologies and solutions to enhance surveillance and response systems for emerging zoonotic diseases. Sci One Health. 2024;3:100061. doi:10.1016/j.soh.2023.100061

9. Zahlan A, Ranjan R, Hayes D. Artificial intelligence innovation in healthcare: literature review, exploratory analysis, and future research. Technology in Society. 2023;74: 102321. doi:10.1016/j.techsoc.2023.102321.

10. Division of Epidemiology, Department of Disease Control. Guidelines for reporting communicable diseases under surveillance: Digital format in accordance with the Communicable Disease Act B.E. 2558. Paper presented at: Meeting of the guidelines for submitting epidemiological surveillance data in digital format (D506); 2023 Sep 21; Bangkok, Thailand. Thai.

11. Division of Epidemiology, Department of Disease Control. Case definition for Communicable Diseases Surveillance, Thailand, 2020. Nonthaburi: Division of Epidemiology, Department of Control (TH); 2020. Thai.

12. Wongsanuphat S, Sangwanloy S, Sopha P, Buangsuang W, Denduang S, Thongplean A, et al. Enhancingcoronavirus disease (COVID-19) surveillance system through information technology, Thailand, 2020. OSIR. 2020 Sep;13(3):90–100.

13. Wongsanuphat S, Jitpeera C, Iamsirithaworn S, Laosiritaworn Y, Thammawijaya P. An Evaluation of the Enhanced Information System for COVID-19 Surveillance in Thailand, 2020: A Pre-Post Intervention Comparison. OSIR. 2020 Sep;13(3):101–9. doi:10.59096/osir.v13i3.262806.

14. Laudon KC, Laudon JP. Management information systems: managing the digital firm. 16th ed. London: Pearson; 2021.

15. Gorton I. Essential software architecture. 2nd ed. Heidelberg: Springer; 2011.

16. Provost F, Fawcett T. Data science for business. Sebastopol (CA): O'Reilly Media; 2013.

17. Russell S, Norvig P. Artificial intelligence: a modern approach. 4th ed. London: Pearson; 2021.

18. Malaikham J, Suriya S, Prommongkhol J, Wongsuwanphon S, Wongsanuphat S. Development of innovations for analyzing surveillance situations of communicable diseases in Thailand using the design thinking process. Weekly Epidemiological Surveillance Report. 2025;56(3): 1–15. doi:10.59096/wesr.v56i3.3571.

19. Yotwattana P, Tepsittha K. The development of a digital epidemiological surveillance reporting platform. Weekly Epidemiological Surveillance Report. 2023;54:717–30.

20. Centers for Disease Control and Prevention (US). BioSense platform [Internet]. Atlanta: Centers for Disease Control and Prevenion: 2024 Apr 22 [cited 2024 Nov 4]. <https://www.cdc.gov/nssp/php/about/about-nssp-and-the-biosense-platform.html>

21. Doig K, Papenfuss AT, Fox S. Clinical cancer genomic analysis: data engineering required. Lancet Oncol 2015 Sep;16(9):1015–7. doi:10.1016/s1470-2045(15)00195-3.

22. Cui C, Chou SS, Brattain L, Lehman CD, Samir AE. Data engineering for machine learning in women's imaging and beyond. AJR Am J Roentgenol. 2019 Jul;213(1):216–26. doi:10.2214/ajr.18.20464.

23. Zhang Q. Data science approaches to infectious disease surveillance. Philos Trans A Math Phys Eng Sci. 2021 Nov 22;380(2214): 20210115. doi:10.1098/rsta.2021.0115.

24. Hanan M, Galal GEH. Design thinking using qualitative data analysis and machine learning. Proceedings of the 2023 13th International Conference on Information Communication and Management. 2023 Nov 7; Cairo, Egypt. New York: Association for Computing Machinery; 2024. p. 40–7. doi:10.1145/3640429.3640437.

25. Nguyen AH, Hoang TG, Nguyen LQ, Thi Pham HM. Design thinking-based data analytic lifecycle for improving management control in Banks. Technology Analysis & Strategic Management. 2022 Jul 14;36(7):1508–23. doi:10.1080/09537325.2022.2100754.

26. Drzyzga G, Harder T. User-centered design and iterative refinement: Promoting student learning with an interactive dashboard. Proceedings of the 19th International Conference on Web Information Systems and Technologies. 2023 Nov 15–17; Rome, Italy. Setubal (PT): Science and Technology; p. 340–6. doi:10.5220/0012191300003584

27. Lennox-Chhugani N. A user-centred design approach to integrated information systems – a perspective. Int J Integr Care. 2018;18(2):15. doi:10.5334/ijic.4182.

28. Areechokchai D, Sayumpurujinan S. Developing data visualization for hospital-based surveillance under the Communicable Disease Act B.E. 2558. Weekly Epidemiological Surveillance Report. 2023;55(9):1–12. doi:10.59096/wesr.v55i9.3247.

29. Althomsons SP, Winglee K, Heilig CM, Talarico S, Silk B, Wortham J, et al. Using machine learning techniques and national tuberculosis surveillance data to predict excess growth in genotyped tuberculosis clusters. Am J Epidemiol. 2022;191(11):1936–43. doi:10.1093/aje/kwac117.

30. Panja M, Chakraborty T, Kumar U, Liu N. Epicasting: an ensemble wavelet neural network for forecasting epidemics. Neural Netw. 2023; 165:185–212. doi:10.1016/j.neunet.2023.05.049.

31. Didi Y, Walha A, Wali A. Data-driven model for influenza prediction incorporating environmental effects. Proceedings of the 5th International Conference on Internet of Things, Big Data and Security; 2020 May 7–9; Prague, Czech Republic. Setubal (PT): Science and Technology; 2022. p. 15–24. doi:10.5220/0009325500150024.

32. Shinde S, Yadav S, Somvanshi A. Epidemic outbreak prediction using machine learning model. Proceedings of the 5th International Conference on Advances in Science and Technology (ICAST); 2022 Dec 2-3, Mumbai, India. New York: IEEE; 2023. p. 127–32. doi:10.1109/ICAST55766.2022.10039594.

33. El Morr C, Ozdemir D, Asdaah Y, Saab A, El-Lahib Y, Sokhn ES. Ai-based epidemic and pandemic early warning systems: a systematic scoping review. Health Informatics J. 2024 Jul-Sep;30(3):14604582 241275844. doi:10.1177/14604582241275844.

34. Ahn E, Liu N, Parekh T, Patel R, Baldacchino T, Mullavey T, et al. A mobile app and dashboard for early detection of infectious disease outbreaks: development study. JMIR Public Health Surveill 2021;7(3):e14837. doi:10.2196/14837.

35. Mohanty B, Chughtai A, Rabhi F. Use of mobile apps for epidemic surveillance and response – availability and gaps. Global Biosecurity. 2019 Sep 11;1(2):37. doi:10.31646/ gbio.39.

36. Sun S, Shaw M, Moodie EEM, Ruths D. The epidemiological impact of the Canadian covid alert app. Can J Public Health. 2022;113(4):519–27. doi:10.17269/s41997-022-00632-w.

37. Python Software Foundation. Python: a programming language [Internet]. Version 3.13.2. Beaverton (OR): Python Software Foundation; [cited 2025 Mar 16]. <https://www.python.org/>

38. Tableau Software. Tableau Desktop [Internet]. Version 2024.3.4. Seattle (WA): Salesforce; [cited 2025 Mar 16]. <https://www.tableau.com/>