Review of Contemporary Philosophy
ISSN: 1841-5261, e-ISSN: 2471-089X
Vol 22 (1), 2023
Pp 2926 - 2935

Advancements in Syndromic Surveillance Systems: A Comprehensive Review

[1]-Salman Ibrahim N Aljadid,[2]- Faisal Ahmed Ahmed Rostom,[3]- Khalid Saud Alanazi,[4]- Faisal Razhan Muhammad Al-Shehri,[5]- Adel Hassan Yahya Mashufi,[6]- Mansour Adam Hamid Farag,[7]- Mohammed Salem Hadhidh Alhejaili,[8]- Abdulelah Abdullahalmansour,[9]-Mohammed Omar Mohammed Alsherif,[10]- Nawar Dakeel Allah Safer Alotaibe,11-Hassan Mubarak Saqer,[12]_ Abdulmohsen Abdulaziz Alaiban,[13]-Ali Essa Ahmed Dallak,[14]-Awaad Asaad Ali Alwadani,[15]-Mousa Ahmed Ali Sahli,[16]-Yahya Mousa Mohammed Sulayyi

1. Ksa, Ministry Of Health, Prince Sultan Military Medical City
2. Ksa, Ministry Of Health, Primary Health Care Center In The Southern Rawda District, Jizan
3. Ksa, Ministry Of Health, Public Health Specialist At Al-Mruoj Center In Tabuk
4. Ksa, Ministry Of Health, Al Rawdah Health Center 1, Riyadh Health
5. Ksa, Ministry Of Health, Sabya Hospital
6. Ksa, Ministry Of Health, Primary Care Sector
7. Ksa, Ministry Of Health, Prince Mohammed Bin Abdulaziz Airport Medina
8. Ksa, Ministry Of Health, Hosptal Lmamabdulrahmanalfaisal-Riyadh
9. Ksa, Ministry Of Health, Al-Wajh General Hospital
10. Ksa, Ministry Of Health, Arwa Primary Health Care Center
11. Ksa, Ministry Of Health, Al Farsha General Hospital
12. Ksa, Ministry Of Health,
13. Ksa, Ministry Of Health, Sabya Hospital
14. Ksa, Ministry Of Health, Jazan Health Cluster
15. Ksa, Ministry Of Health, Abuarish General Hospital
16. Ksa, Ministry Of Health, Al-Madha Genaral Hospital-Asir

Abstract

Background: Syndromic surveillance systems have emerged as critical tools for monitoring communicable infectious diseases, leveraging data from diverse sources, including social media platforms. This review examines recent advances in syndromic surveillance, focusing on the integration of real-time data and machine learning techniques to enhance disease detection and public health response.

Methods: A comprehensive literature review was conducted, analyzing publications from 2010 to 2023 found in databases such as ACM Portal, IEEE Xplore, Science Direct, and PubMed.

Results: The findings reveal that syndromic surveillance systems increasingly utilize social media data to identify disease outbreaks faster than traditional methods, which often rely on clinical diagnoses that can introduce significant delays. Machine learning algorithms, particularly natural language processing, have improved the accuracy of data analysis and disease prediction. Results indicate that Twitter is the most frequently utilized platform for health-related data collection, and supervised machine learning approaches, especially the Support Vector Machine (SVM), demonstrate superior performance in classifying health-related tweets. This review highlights the effectiveness of syndromic surveillance systems in providing early warnings for various health threats, including influenza and other infectious diseases.

Conclusion: The study concludes that the integration of social media and machine learning into syndromic surveillance represents a promising evolution in public health monitoring, enabling quicker and more efficient responses to emerging health threats. However, challenges such as data noise, misinformation, and privacy concerns remain significant barriers to the effective implementation of these systems.

Keywords: Syndromic surveillance, social media, Machine learning, Public health, Disease detection.

Received: 10 october 2023 Revised: 24 November 2023 Accepted: 08 December 2023

1. Introduction

Syndromic surveillance systems are designed to gather data that facilitates an overview of all communicable infectious illnesses. These systems may differ based on the data source, intended length, and methods of data recording and acquisition. These systems may use conventional data and real-time information from diverse social media sites. In healthcare, these systems often emphasize the early detection of sickness clusters and the symptomatic phase before to the confirmation of a specific disease by any clinical unit or laboratory, facilitating a swift reaction. Surveillance systems often focus on the systematic collecting, analysis, and interpretation of data, as well as the detection, confirmation, and reporting of diseases, while also addressing public health responses. Objectified definitions, algorithmic diagnostics, and electronic databases have enhanced the user-friendliness and efficacy of surveillance systems throughout time [1]. Conventional biosurveillance depends on clinical interactions to gather data, a method that is labor-intensive. Furthermore, conventional pandemic monitoring mostly relies on manual procedures, resulting in a delay of one to two weeks in the accessibility of data derived from clinical diagnoses [2,3].

In recent years, the emergence of web-based data sources has augmented conventional surveillance methods and has consistently enhanced infectious disease monitoring by offering real-time information and lowering public health costs. It is noteworthy how swiftly occurrences may be identified in real-time with internet-based surveillance data, since an average individual's social media post has resulted in a significant surge in public participation with skin cancer prevention [6]. In addition to its many applications, the function of monitoring social media warrants examination in the context of healthcare decision-making. The early identification of illnesses and prompt public health interventions have necessitated the development of novel methodologies and technologies to enhance the efficacy of conventional syndromic monitoring systems. Numerous research papers and articles have emerged that analyze different surveillance systems using social media data [5,9-18]. These publications have addressed the applications, technologies, algorithms, data sources, and their assessment. To the best of our knowledge, a current study of surveillance systems in the health informatics sector using social media is unavailable. This article reviews recent machine learning technologies and methodologies, highlighting problems and prospects in the field.

2. Method

This research examines social media-based monitoring systems that use machine learning techniques to forecast diseases in real-time or near real-time. The criterion for selecting research publications were those published between 2010 and 2023. The following scientific databases were examined to compile a complete bibliography of research publications concerning social media-based surveillance systems in the healthcare sector: ACM Portal, IEEE Xplore, Science Direct, and PubMed.

3. Machine learning techniques used by surveillance systems for analyzing social media data

In recent years, machine learning has garnered significant interest, particularly in the analysis of patterns within pictures or raw data. Bates [19] discussed how advancements in machine learning enable epidemiologists to analyze extensive digital datasets. Mike and Daniel [11] examined the integration of natural language processing and machine learning with social media platforms to facilitate the study of large datasets for population-level mental health research. Among several methodological variants of machine learning, some architectures exhibit notable appeal. We observed that the accuracy of the k-Nearest Neighbor (k-NN) classifier surpassed that of many other machines learning classifiers, including

NBM Modal, NB, and SVM [20], in categorizing tweets into two classes: actual instances of allergies or awareness tweets.

Lee et al. [21] demonstrated that optimal text classification performance was achieved using the Multinomial Naive Bayes model, yielding an F-measure of 0. 811 in comparison to other classifiers, including Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). The author [22] demonstrated a model using a multilayer perceptron with a backpropagation algorithm on Twitter data to forecast the weekly incidence of ILI among the US population. Various supervised machine learning algorithms were examined for identifying personal health experience tweets and used the deep gramulator method to enhance accuracy when tested on separate datasets [23].

Unsupervised classification algorithms do not need labeled datasets for output prediction, unlike supervised algorithms. Consequently, unsupervised classification approaches seem to be a more appealing option for text analysis; nonetheless, they may pose more challenges in attaining comparable accuracy to supervised methods. Similar observations may be seen when Sousa et al. [24] conducted the categorization of tweets using both supervised and unsupervised methodologies. They found that topic modeling (LDA), an unsupervised approach, offers less control over topic material than a standard classifier, especially in a naturally noisy media channel. Consequently, Multinomial Naive Bayes, a supervised classification method, was used for categorizing Twitter material. The extensive use of machine learning algorithms for the analysis and differentiation of health-related social media data encompasses [20,21,23,25].

4. Prevalent social media data repositories for data acquisition

Due to a substantial rise in social media users sharing information, researchers are more motivated to examine social media activities for public health objectives. Furthermore, social media may include a range of subjects beyond those addressed by conventional public health data sources. Social media has become a viable medium for health communication [10,26-28]. While several studies have questioned the efficacy of social media data in epidemic detection, the analysis of social media material for healthcare information has garnered significant attention [29-33]. To monitor and predict health occurrences, [34] emphasizes the critical need of using social media data sources when sending epidemic outbreak alerts via Facebook or Twitter to reach a broader public audience promptly. Consequently, social media postings and internet search patterns may serve as valuable sources of information on health epidemics.

Twitter is a prominent micro-blogging platform that allows registered users to submit tweets or retweet content, which may be accessed by unregistered users. Twitter, with over 300 million monthly active users, has emerged as a dependable and rapid platform for assessing illness prevalence across a community. Kwak et al. [35] researched to investigate the potential of Twitter as a novel information-sharing platform. Given that social media posts from Twitter have emerged as a dependable and rapid means to assess illness prevalence among a community, it is imperative to devise effective and efficient methodologies for processing and analyzing health-related tweets. Typically, factors such as location, volume, time, and public perceptions are taken into account for disease monitoring.

A recent study [14] used data gathered from Twitter to extract insights throughout several outbreaks that proved beneficial for many health organizations. Bosley et al. used a set of seven phrases to aggregate over 60,000 tweets, then analyzed and categorized them with an emphasis on cardiac arrest and resuscitation [36]. To facilitate the early identification of influenza activity, [37] suggested a model using real-time Twitter data streams and U.S. information. Historical statistics from the Centers for Disease Control and Prevention (CDC) to predict future influenza activity. Lee et al. [21] investigated the activities related to allergies. They gathered tweets that referenced allergies. A natural language processing methodology is used [38] to analyze Ebola-related tweets, focusing on four primary issues derived from clusters of keywords: risk factors, preventative education, illness patterns, and compassion. An unsupervised technique was used [13] to categorize vector-represented tweets into clusters of analogous terms. The tweet is thereafter categorized as either disease-related or unrelated according to the similarity metrics of the clusters.

El-bathy et al. [39] suggested a surveillance lifecycle architecture using an innovative evolutionary algorithm to efficiently extract pertinent data from large online datasets at reduced costs. Health issues, including respiratory, gastrointestinal, and heat-related illnesses, and influenza-like disease symptoms, were identified among the populace during large gatherings via Twitter [40]. Twitter data has been utilized to investigate various public health incidents, including allergies, mosquito-borne diseases, dengue, influenza, H1N1, and other diseases.

Instagram is a picture and video-sharing platform established in 2010, boasting over 800 million registered users. Guidry et al. [41] analyzed Ebola-related social media postings on Twitter and Instagram, revealing that Instagram may serve as an effective medium for communication and public engagement during global health emergencies. Research has also investigated Zika-related posts on Instagram [42]. Instagram, as a platform for sharing photos and videos, offers picture and video data that may serve as a valuable resource for disease monitoring.

Crowdsourcing is a method whereby a substantial, unspecified collective of volunteers or part-time workers contributes services, ideas, or material via a flexible open invitation [43]. A diverse assembly of individuals with differing levels of expertise and experience often collaborates towards a shared objective. Prominent crowdsourcing approaches include Google Consumer Surveys, Amazon Mechanical Turk, and specialized websites. EO et al. [44] examined the influence of reviews on food service on Yelp.com, a business review platform, on foodborne disease surveillance initiatives, and noted that online illness reporting might enhance conventional surveillance methods. Lwin et al. [45] have aggregated crowdsourced data on mosquito bites, symptoms, and potential breeding places, subsequently reporting this information to assist health authorities in providing early warnings of dengue epidemics. A solution that integrates crowdsourcing with text classification techniques was developed to monitor disinformation about the Zika virus on Twitter [46,47].

5. Disease prediction based on syndromic surveillance

Syndromic surveillance has developed into a viable instrument for forecasting epidemics for public health objectives by using data collected from diverse sources before the availability of clinically validated information. The objective is to reduce the dispersion throughout the population and implement preventative measures. In recent years, social media data has been extensively used to assess illness prevalence and identify disease outbreaks. This data might aid public health professionals in the early identification of epidemics compared to conventional approaches. The data typically consists of self-reported symptoms. Research indicates that monitoring systems in healthcare can forecast illnesses relevant to public health issues. Twitter data served as a mechanism for early warning and epidemic identification, facilitating predictions for syphilis, swine flu, TB, influenza, and Ebola [48,49]. Research conducted in the Philippines examined the prevalence of dengue and typhoid illness in a specific area [29]. Additionally, another research [41] indicates that monitoring systems have used social media data for illness detection.

6. Assessment of illness magnitude during a certain duration

Surveillance systems may be used to assess the extent of the issue. Planning, resource allocation, therapies, and prevention may be executed by forecasting future illness levels. The analysis from monitoring systems is valuable for assessing the disease's progression over time, allowing for informed evaluations. Event-based surveillance is the rapid and systematic collection of data on incidents that pose potential risks to public health. The data may originate from several Internet sources, including media reports, online discussion forums, regular reporting systems, personal information, or hearsay. In the context of an online forum, an event is defined as an overabundance of news submissions. The importance of the event is commensurate with the volume of discussions around it.

Consequently, the impact of the event may be seen in subject dispersion by the volume of postings related to the issue. When the volume of posts on a subject exceeds the typical amount, it is reasonable to anticipate that an event is occurring at that moment. Notable instances of event-based monitoring systems include

HealthMap, which monitors health-related news, and EpiSPIDER, among others. The use of social media for public health intelligence monitoring to enhance situational awareness of possible health concerns and assist surveillance efforts has significantly increased in recent decades [14]. The research examined the Zika virus pandemic using a Twitter corpus. Thapen et al. [50] introduced DEFENDER, a software system capable of detecting possible health incidents by monitoring Twitter streams and then presenting the created events to users via a Front-end UI.

Surveillance systems in healthcare may assess social media users' responses to health promotion messages or events. Tran and Lee [51] collected 2 billion geotagged tweets in 90 languages from Twitter between August 2014 and December 2014 to analyze civilian responses to Ebola. Further research gathered Twitter data to convey feelings during various phases of a public health emergency [50]. Choi et al. [52] examined the correlation between mass media and emotional public responses during the nationwide MERS epidemic in Korea in 2015. Seltzer et al. [42] posited that public opinions might be articulated via Instagram and highlighted issues pertinent to public health. Further research [53] indicated the inclination of Twitter users to disseminate information on the Zika virus, including symptoms, narratives of pregnant women afflicted with Zika, and concerns of expectant parents.

Furthermore, Gaspar et al. [54] showed that the amplification of public responses on Twitter during the 2011 EHEC/*Escherichia coli* epidemic in Europe had significant political and economic repercussions. Liu et al. examined variances in reactions to various disease outbreaks by investigating public search behavior [55]. In a separate investigation, Fung et al. analyzed social media data to understand Chinese individuals' responses to various epidemics [56].

Upon detection of the incident, surveillance technologies may be used to assess the public's knowledge and perception of health events. Sentiments expressed by users on social media in an epidemic scenario indicate their knowledge, attitudes, and perceptions. Social media facilitate the dissemination of public attitudes, ideas, and reactions during epidemics [57]. The significant uncertainty surrounding Zika may drive users to social media for crucial information and has also contributed to the establishment and dissemination of conspiracy theories [58,59].

7. Monitoring misinformation

The capacity of social media to disseminate information in real time also facilitates the rapid spread of disinformation among individuals during an ongoing pandemic, which might elicit serious consequences. Nonetheless, to our knowledge, there has been no study on the detection of disinformation about such events, with just a few academics addressing the pervasive misinformation disseminated on social media platforms. Ghenai and Mejova [46] used machine learning methodologies to monitor disinformation about the Zika infection on the Twitter network. A separate research [12] investigated methods to rectify misconceptions about the measles vaccine by analyzing the responses of two social media groups. Subsequently, they investigated the impact of rectifying the misinformation on users who responded to the measles vaccine. Research [60] investigated fraudulent medical news on social media. Further research using YouTube data revealed that the dissemination of both factual content and conspiracy theories occurred comparably, which was a significant finding for health organizations. To combat the proliferation of disinformation, more scrutiny is necessary about the material disseminated on YouTube and other internet platforms. Separate research examined 1,680 microblog postings and found that over 20% included incorrect information. To address the deficiency in how health organizations should elicit responses and rectify misinformation, Hagg et al. [32] indicated that the risk of misrepresentation may impede the use of social media data in healthcare.

8. Constraints and obstacles of monitoring systems based on social media

An issue encountered during data collection is noise. The data gathered from social media platforms may include information that is extraneous to the goal. Data about sickness terminology has little correlation to health. A significant volume of flu-related activity may stem from postings including the keyword “Irish Flu.” Regrettably, users may sometimes post a status indicating illness, leading to unfounded suspicions of infection. Thus, misleading information might affect illness management within the public health department. To mitigate data noise, fundamental text processing techniques like as feature weighting, tokenization, stemming, and frequency-based algorithms are available; nevertheless, further training is necessary to get useful data for further analysis.

A further problem related to social media data is its verification. The use of unauthorized data from social media may result in issues related to standardization, verification, and control [61]. Bodnar and Salathé concentrated on verifying the ground truth linked to a substantial volume of diverse social media data [62]. The dataset is a fundamental element in the development of predictive models. The efficacy of the prediction models is significantly contingent upon the dataset. The dataset for the predictive models comprises historical information, training data, and testing data. A substantial volume of training data is necessary to develop forecasting models, together with testing data for the valid evaluation of predictions derived from the model training. Therefore, the accuracy of data is equally crucial as the quantity and quality of data from trustworthy sources [63].

A further difficulty associated with social media-derived statistics is diminished trust. Research [64] indicates that government websites are a more reliable source of vaccination information than social media. The quality of online health data is inconsistent. Numerous social media data analysis techniques may reveal spikes indicative of significant occurrences; however, they may represent fear rather than actual instances of a disease epidemic. Additionally, users may claim to have influenza when they have a normal cold, or individuals could discuss illnesses due to heightened media sensationalism. The research discusses conspiracy theories about the Zika virus epidemic on Reddit during a public health catastrophe [65]. Such activities contribute to the diminished faith in internet data. Additional training is necessary to adjust the categorization algorithms to avert this problem.

While data such as location, username, follower count, and user profiles can be aggregated alongside tweet content, demographic details like age, gender, and race are challenging to ascertain from tweets, complicating the identification of individuals discussing the outbreak and the targeting of public health initiatives. Among the few research that has examined user profiles, it was observed that males generated a higher volume of Facebook comments per individual compared to women. Another bias is that younger individuals constitute the predominant demographic of social media users. The bias is significantly corroborated by the finding that social media statistics are mostly derived from active users, who are often young people, well-educated, and possess substantial money. This article used machine learning approaches to demonstrate that sociodemographic characteristics significantly influence the reporting of disease-related postings on social media. Consequently, social media users are not deemed representative of the whole population [66]. Any result derived from social media data omits those who do not use these platforms, likely including the most susceptible segments of the population, as well as those who often choose not to disclose their health issues publicly. It may also be inferred that those experiencing pain, illness, advanced age, or disabilities are less likely to be active users.

Language constitutes an additional bias influenced by the demography of a group. Many studies concentrate on a single language (English), even though the core of the epidemic/event is located in a nation where English is not an official language [67]. Limited studies have shown research using languages other than English for illness monitoring. For example, Ilyas [68] has spatially categorized illness names in tweets in both Arabic and English. Furthermore, the study noted that tweets in English constituted a minimal proportion. Their methodology was validated by a strong association between the actual incidence of MERS-Coronavirus illnesses and disease-related tweets in Arabic.

There is an ongoing discourse on ethical considerations in data retrieval from social media, particularly concerning the privacy of datasets amassed for health-related objectives [69,70]. Despite the public accessibility of social media data, individuals may prefer that their postings or information not be used for research purposes. Hosts of social media platforms must take into account users' privacy expectations, including diagnoses derived from public data and apprehensions around algorithms that disclose implicit user demographics. Only a limited number of research [71,72] have shown interest so far. Consequently, there is an expanded opportunity to contemplate privacy while researching social media data.

While communication via social media facilitates the retrieval of healthcare information, it poses challenges with the semantic interpretation of language. Social media writings are characterized by informality and ambiguity; hence, just matching keywords may fail to capture their semantic meaning, leading to incomplete results. Limsopatham and Collier [73] have examined and investigated this problem.

9. Conclusions

Among all the social media sites examined in the chosen articles, Twitter (64%) emerged as the most frequently searched medium. Furthermore, SVM (33%) was the predominant classification approach among all recently used machine learning classifiers. Furthermore, it was noted that SVM emerged as the best effective classifier for binary classification tasks. This research intends to examine contemporary developments in surveillance systems that use social media and machine learning algorithms within the domain of public health. We have observed that social media-based surveillance systems exhibit advantages over conventional surveillance methods. We have also examined the use of dynamic social media data to enhance monitoring systems in public health. Furthermore, the issues outlined in this work are anticipated to catalyze the development of novel technologies and enhanced capacities for public health research.

References

1. Mathur P. Need for National-level Surveillance of HAIs in India. Journal of Clinical Infectious Diseases Society. 2023 Apr 1;1(2):137-9.
2. K. Lee, Real-Time Disease Surveillance Using Twitter Data : Demonstration on Flu and Cancer, (n.d.) 1474–1477.
3. R.W. Newkirk, J.B. Bender, C.W. Hedberg, The Potential Capability of Social
4. Media as a Component of Food Safety and Food Terrorism Surveillance Systems, Foodborne Pathog. Dis. 9 (2012) 120–124.
5. D.B. Neill, New directions in artificial intelligence for public health surveillance, IEEE Intell. Syst. 27 (2012) 56–59.
6. S.J. Yan, A.A. Chughtai, C.R. Macintyre, Utility and potential of rapid epidemic intelligence from internet-based sources, Int. J. Infect. Dis. 63 (2017) 77–87.
7. S.M. Noar, E. Leas, B.M. Althouse, M. Dredze, D. Kelley, J.W. Ayers, Can a selfie promote public engagement with skin cancer ? Prev. Med. (Baltim). (2017) 1–4.
8. I.C.-H. Fung, Z.T.H. Tse, K.-W. Fu, The use of social media in public health surveillance, 6 (2015) 10–13.
9. L. Mollema, I.A. Harmsen, E. Broekhuizen, R. Clijnk, H. De Melker, T. Paulussen, G. Kok, R. Ruiter, E. Das, Disease Detection or Public Opinion Reflection ? Content Analysis of Tweets , Other Social Media , and Online Newspapers During the Measles Outbreak in the Netherlands in 2013, 17 (2013) 1–12.
10. S.J. Park, S. Hong, D. Kim, Y. Seo, J.H. Hur, W. Jin, D. Precision, Development of a Real-Time Stroke Detection System for Elderly Drivers Using Quad-Chamber Air Cushion and IoT Devices, (2018) 1–5.
11. I.C.H. Fung, C.H. Duke, K.C. Finch, K.R. Snook, P.L. Tseng, A.C. Hernandez, M. Gambhir, K.W. Fu, Z.T.H. Tse, Ebola virus disease and social media: A systematic review, Am. J. Infect. Control. 44 (2016) 1660–1671.
12. A. Mike, C. Daniel, Social media, big data, and mental health: current advances and ethical implications, Curr. Opin. Psychol. (2016).
13. L. Tang, B. Bie, S.P. Ma, D. Zhi, Social media and outbreaks of emerging infectious diseases: A systematic review of literature, AJIC Am. J. Infect. Control. (2018).
14. X. Dai, M. Bikdash, B. Meyer, From Social Media to Public Health Surveillance: Word Embedding based Clustering Method for Twitter Classification, 2017.
15. J. O'Shea, Digital disease detection: A systematic review of event-based internet biosurveillance systems, Int. J. Med. Inform. 101 (2017) 15–22.

16. L. Fernandez-luque, M. Imran, Humanitarian Health Computing using Artificial Intelligence and Social Media: A Narrative Literature Review, Int. J. Med. Inform. (2018).
17. A. Alessa, M. Faezipour, A review of influenza detection and prediction through social networking sites, (2018) 1–27.
18. H.A. Park, H. Jung, J. On, S.K. Park, H. Kang, Digital epidemiology: Use of digital data collected for non-epidemiological purposes in epidemiological studies, Healthc. Inform. Res. (2018).
19. T. Eckmanns, K. Denecke, E. Velasco, T. Agheneza, G. Kirchner, Social Media and Internet-Based Data in Global Systems for Public Health Surveillance: A Systematic Review, Milbank Q. 92 (2014) 7–33.
20. M. Bates, Tracking Disease: Digital Epidemiology Offers New Promise in Predicting Outbreaks, IEEE Pulse. 8 (2017) 18–22
21. K. Nargund, S. Natarajan, Public health allergy surveillance using micro-blogs, in: 2016 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2016. (2016) 1429–1433.
22. K. Lee, A. Agrawal, A. Choudhary, Datasets, Mining Social Media Streams to Improve Public Health Allergy Surveillance, (2015) 815–822.
23. K. Lee, A. Agrawal, A. Choudhary, Forecasting Influenza Levels Using Real-Time Social Media Streams, in: Proc. - 2017 IEEE Int. Conf. Healthc. Informatics, ICHI 2017. (2017) 409–414.
24. R.A. CaliX, R. Gupta, M. Gupta, K. Jiang, Deep gramulator: Improving precision in the classification of personal health-experience tweets with deep learning, in: Proc. - 2017 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2017. 2017-Janua (2017) 1154–1159.
25. L. Sousa, R. de Mello, D. Cedrim, A. Garcia, P. Missier, A. Uchôa, A. Oliveira, A. Romanovsky, VazaDengue: An information system for preventing and combating mosquito-borne diseases with social networks, Inf. Syst. 75 (2018) 26–42.
26. W.D. Jenkins, B. Wold, Use of the Internet for the surveillance and prevention of sexually transmitted diseases, Microbes Infect. 14 (2012) 427–437.
27. G.D. Haddow, K.S. Haddow, G.D. Haddow, K.S. Haddow, Chapter Eleven –Communicating During a Public Health Crisis, Disaster Commun. a Chang. Media World. (2014) 195–209.
28. K. Denecke, P. Dolog, P. Smrz, Making Use of Social Media Data in Public Health, (2012) 243–246.
29. E. Yom-tov, Ebola data from the Internet: An Opportunity for Syndromic Surveillance or a News Event? Categories and Subject Descriptors, (n.d.) 115–119.
30. T. Nguyen, M.E. Larsen, B.O. Dea, D.T. Nguyen, J. Yearwood, D. Phung, S. Venkatesh, H. Christensen, Kernel-based features for predicting population health indices from geocoded social media data, (2017).
31. P. Kostkova, A Roadmap to Integrated Digital Public Health Surveillance : the Vision and the Challenges, (2013) 687–693.
32. E. Hagg, V.S. Dahinten, L.M. Currie, The emerging use of social media for health-related purposes in low and middle-income countries: A scoping review, Int. J. Med. Inform. 115 (2018) 92–105.
33. T.H. Van De Belt, P.T. Van Stockum, L.J.L.P.G. Engelen, J. Lancee, R. Schrijver, J. Rodríguez-baño, E. Tacconelli, K. Saris, M.M.H.J. Van Gelder, A. Voss, Social media posts and online search behaviour as early-warning system for MRSA out- breaks, (2018) 1–10.
34. S. Chaudhary, S. Naaz, Use of Big Data in Computational Epidemiology for Public Health Surveillance, (2017) 150–155.
35. H. Kwak, C. Lee, H. Park, S. Moon, What is Twitter, a Social Network or a News Media? Arch. Zootec. 60 (2011) 297–300.
36. A. Stefanidis, E. Vraga, G. Lamprianidis, J. Radzikowski, P.L. Delamater, K.H. Jacobsen, D. Pfoser, A. Croitoru, A. Crooks, Zika in Twitter: Temporal Variations of Locations , Actors , and Concepts Corresponding Author:, 3 (n.d.). 10.2196/ publichealth.6925.
37. J.C. Bosley, N.W. Zhao, S. Hill, F.S. Shofer, D.A. Asch, L.B. Becker, R.M. Merchant, Decoding twitter: Surveillance and trends for cardiac arrest and resuscitation communication, Resuscitation 84 (2013) 206–212.
38. M.O. Edd, S.Y. Rn, What can we learn about the Ebola outbreak from tweets ? Am.J. Infect. Control. 43 (2015) 563–571.

39. N. El-bathy, C. Gloster, M. El-bathy, G. Stein, R. Stevenson, Intelligent Surveillance Lifecycle Architecture for Epidemiological Data Clustering using Twitter and Novel Genetic Algorithm, (2014) 149–155.
40. Arjona JO. Redes sociales geolocalizadas y COVID-19: análisis de la actividad espaciotemporal de los usuarios de Twitter de España durante la pandemia. GeoFocus. International Review of Geographical Information Science and Technology. (2022) Dec 30(30):25-47.
41. J.P.D. Guidry, Y. Jin, C.A. Orr, M. Messner, S. Meganck, Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement, Public Relat. Rev. 43 (2017) 477–486.
42. E.K. Seltzer, E. Horst-Martz, M. Lu, R.M. Merchant, Public sentiment and discourse about Zika virus on Instagram, Public Health. 150 (2017) 170–175.
43. E.E. Arolas, F.G. Ladrón-de-Guevara, Towards an integrating crowdsourcing definition, 32 (2016) 189–200.
44. N. EO., K. SA., B. JS., Online reports of foodborne illness capture foods implicated in official foodborne outbreak reports., Prev. Med. (Baltim). 67 (2014) 264–269.
45. M.O. Lwin, S. Vijaykumar, O.N.N. Fernando, S.A. Cheong, V.S. Rathnayake, G. Lim, Y.L. Theng, S. Chaudhuri, S. Foo, A 21st century approach to tackling dengue: Crowdsourced surveillance, predictive mapping and tailored communication, Acta Trop. 130 (2014) 100–107.
46. A. Ghenai, Y. Mejova, Catching Zika Fever : Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter, (2017) 5090.
47. P. Quade, E.O. Nsoesie, P. Quade, A Platform for Crowdsourced Foodborne Illness Surveillance : Description of Users and Reports, 3 (2017) 1–9.
48. X. Zhou, J. Ye, Y. Feng, Tuberculosis surveillance by analyzing google trends, IEEE Trans. Biomed. Eng. 58 (2011) 2247–2254.
49. S.D. Young, N. Mercer, R.E. Weiss, E.A. Torrone, S.O. Aral, Using social media as a tool to predict syphilis, Prev. Med. (Baltim) 109 (2018) 58–61.
50. N. Thapen, D. Simmie, C. Hankin, J. Gillard, DEFENDER : Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response, (2016) 1–19.
51. T. Tran, K. Lee, Understanding Citizen Reactions and Ebola-Related Information Propagation on Social Media, (2016) 106–111.
52. S. Choi, J. Lee, M. Kang, H. Min, Y. Chang, S. Yoon, Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks, Methods 129 (2017) 50–59.
53. K.W. Fu, H. Liang, N. Saroha, Z.T.H. Tse, P. Ip, I.C.H. Fung, How people react to Zika virus outbreaks on Twitter? A computational content analysis, Am. J. Infect. Control. 44 (2016) 1700–1702.
54. R. Gaspar, S. Gorjão, B. Seibt, L. Lima, J. Barnett, A. Moss, J. Wills, Tweeting
55. during food crises: A psychosocial analysis of threat coping expressions in Spain, during the 2011 European EHEC outbreak, Int. J. Hum. Comput. Stud. 72 (2014) 239–254.
56. K. Liu, L. Li, T. Jiang, B. Chen, Z. Jiang, Z. Wang, Y. Chen, J. Jiang, H. Gu, Chinese public attention to the outbreak of ebola in west africa: Evidence from the online big data platform, Int. J. Environ. Res. Public Health. (2016).
57. I.C. Fung, K. Fu, Y. Ying, B. Schaible, Y. Hao, C. Chan, Z.T.-H. Tse, Chinese social media reaction to the MERS-CoV and avian influenza A (H7N9) outbreaks, 2013.
58. L. Tang, B. Bie, D. Zhi, Tweeting about measles during stages of an outbreak: A semantic network approach to the framing of an emerging infectious disease, Am.J. Infect. Control. 46 (2018) 1375–1380
59. K.H. Pine, Y. Chen, W. Lafayette, Managing Uncertainty : Using Social Media for Risk Assessment during a Public Health, Crisis (2017).
60. M.H. Purnomo, S. Sumpeno, E.I. Setiawan, D. Purwitasari, Keynote Speaker II: Biomedical Engineering Research in the Social Network Analysis Era: Stance Classification for Analysis of Hoax Medical News in Social Media, Proc. Comput. Sci. 116 (2017) 3–9.

61. G. Blouin-Genest, A. Miller, The politics of participatory epidemiology: Technologies, social media and influenza surveillance in the US, Heal. Policy Technol. 6 (2017) 192–197.
62. T. Bodnar, M. Salathé, Validating Models for Disease Detection Using Twitter Regression on Tweet Count, (2012) 699–702.
63. A.A. Bharambe, D.R. Kalbande, Techniques and Approaches for Disease Outbreak Prediction, (2016) 100–102.
64. J.R. Cataldi, A.F. Dempsey, S.T. O'Leary, Measles, the media, and MMR: Impact of the 2014–15 measles outbreak, Vaccine. 34 (2016) 6375–6380.
65. Y. Kou, X. Gui, Y. Chen, K. Pine, Conspiracy Talk on Social Media: Collective
66. Sensemaking during a Public Health Crisis, Proc. ACM Human-Computer Interact. 1 (2017) 1–21.
67. L.E. Charles-Smith, T.L. Reynolds, M.A. Cameron, M. Conway, E.H.Y. Lau, J.M. Olsen, J.A. Pavlin, M. Shigematsu, L.C. Streichert, K.J. Suda, C.D. Corley, Using social media for actionable disease surveillance and outbreak management: A systematic literature review, PLoS One. (2015).
68. G. Barata, K. Shores, J.P. Alperin, Local chatter or international buzz? Language differences on posts about Zika research on Twitter and Facebook, PLoS One. (2018).
69. M.U. Ilyas, Disease Tracking in GCC Region Using Arabic Language Tweets, (2018) 417–421.
70. R. Mckee, Ethical issues in using social media for health and health care research, Health Policy (New. York) 110 (2013) 298–301.
71. E.M. Eggleston, E.R. Weitzman, Innovative uses of electronic health records and social media for public health surveillance, Curr. Diab. Rep. (2014).
72. M.A. Mayer, L. Fernández-Luque, A. Leis, Big Data For Health Through Social Media, Elsevier Inc., 2016.
73. N. Limsopatham, N. Collier, Towards the Semantic Interpretation of Personal Health Messages from Social Media, (2015) 27–30.

أنظمة المراقبة السريرية: مراجعة شاملة حول التقدم في دمج وسائل التواصل الاجتماعي وتقنيات التعلم الآلي لتعزيز مراقبة الصحة العامة وكشف الأمراض

الملخص

الخلفية :ظهرت أنظمة المراقبة السريرية كأدوات حيوية لمراقبة الأمراض المعدية القابلة للانتقال، من خلال الاستفادة من البيانات من مصادر متنوعة، بما في ذلك منصات وسائل التواصل الاجتماعي. تستعرض هذه المراجعة التقدمات الأخيرة في المراقبة السريرية، مع التركيز على دمج البيانات في الوقت الحقيقي وتقنيات التعلم الآلي لتعزيز اكتشاف الأمراض واستجابة الصحة العامة.

الطرق :تم إجراء مراجعة شاملة للأدبيات، وتحليل المنشورات من 2010 إلى 2023 الموجودة في قواعد بيانات مثل ACM Portal و IEEE Xplore وScience Direct وPubMed.

النتائج :تكشف النتائج أن أنظمة المراقبة السريرية تستخدم بشكل متزايد بيانات وسائل التواصل الاجتماعي لتحديد تفشي الأمراض بشكل أسرع من الطرق التقليدية، التي غالبًا ما تعتمد على التشخيصات السريرية التي يمكن أن تؤدي إلى تأخيرات كبيرة. لقد حسنت خوارزميات التعلم الآلي، ولا سيما معالجة اللغة الطبيعية، من دقة تحليل البيانات وتوقع الأمراض. تشير النتائج إلى أن تويتر هو المنصة الأكثر استخدامًا لجمع البيانات المتعلقة بالصحة، وأن الأساليب المعتمدة على التعلم الآلي تحت الإشراف، وخاصة آلة الدعم الناقل(SVM) ، تظهر أداءً متفوقًا في تصنيف التغريدات المتعلقة بالصحة. تبرز هذه المراجعة فعالية أنظمة المراقبة السريرية في تقديم تحذيرات مبكرة لمختلف التهديدات الصحية، بما في ذلك الإنفلونزا وغيرها من الأمراض المعدية.

الاستنتاج :يستنتج الدراسة أن دمج وسائل التواصل الاجتماعي والتعلم الآلي في المراقبة السريرية يمثل تطورًا واعدًا في مراقبة الصحة العامة، مما يمكّن من استجابات أسرع وأكثر كفاءة للتهديدات الصحية الناشئة. ومع ذلك، لا تزال التحديات مثل ضوضاء البيانات، والمعلومات المضللة، ومخاوف الخصوصية تشكل حواجز كبيرة أمام التنفيذ الفعال لهذه الأنظمة.

الكلمات المفتاحية :المراقبة السريرية، وسائل التواصل الاجتماعي، التعلم الآلي، الصحة العامة، كشف الأمراض.