

Letter to Editor

Asian Pacific Journal of Tropical Medicine

doi: 10.4103/apjtm.apjtm_194_24

Using X Social Networks (formerly Twitter) and web news mining to predict the measles outbreak

Kia Jahanbin¹, Mohammad Jokar², Vahid Rahmadian³✉¹Department of Computer Engineering Yazd University, Yazd, Iran²Faculty of Veterinary Medicine, Karaj Branch, Islamic Azad University, Karaj, Iran³Department of Public Health, Torbat Jam Faculty of Medical Sciences, Torbat Jam, Iran

Measles, an infectious disease caused by the measles virus, remains a significant public health concern worldwide due to its highly contagious nature and potential for severe complications[1]. In addition to symptoms such as high fever, cough, Koplik spots, and rash, measles can lead to serious complications including pneumonia and myocarditis, particularly in vulnerable populations such as young children[1,2]. In September 2016, the Americas became the first World Health Organization (WHO) region to declare measles eliminated, but intermittent outbreaks continue to occur in developing countries, making it one of the leading causes of childhood death[3]. Despite the availability of a safe and affordable vaccine, it is estimated that there were 128 000 measles-related deaths worldwide in 2021, primarily affecting unvaccinated or inadequately vaccinated children under the age of 5 years, especially in countries characterized by low per capita incomes and inadequate healthcare systems. In 2022, approximately 83% of children worldwide received at least one dose of the measles vaccine by their first birthday through routine health services[1]. A recent report showed that measles also occur in vaccinated individuals, though at reduced severity[3].

With the advent of digital technologies, surveillance and prediction methodologies have revolutionized, which has resulted in the generation of massive real-time data in various fields, including health[2]. These data sources, particularly from X Social Networks (formerly Twitter), web news outlets, and search engines like Google, presents a promising avenue for predicting outbreaks in near real-time. In the context of infectious diseases, prior infodemiological studies have shown the potential utility of social media and Google data analysis[2,4]. Millions of people worldwide use from X Social Networks as a micro-blogging platform to share real-time information, making it a valuable tool for tracking disease-related discussions and identifying potential outbreaks due to its rapid dissemination of news and commentary. Furthermore,

web news mining provides valuable insights into disease trends by automating monitoring, evaluating, and categorizing news articles[5].

One method of controlling and preventing epidemics is to track and analyze social networks regarding the transmission of infectious diseases. In this study, the hybrid deep neural network model of RoBERTa and Bidirectional Gated Recurrent Unit (BiGRU) (HDRB) was employed to analyze the tweets related to measles diseases[6]. Initially, tweets were extracted using X Social Networks Application Programming Interface (API), and preprocessing operations were performed on them. Preprocessing involved standardization of characters, removal of stop words and punctuation, and lemmatization. To implement the HDRB model, Concept Latent Dirichlet Allocation (Concept-LDA) was first utilized to extract various aspects related to measles diseases[7]. Then, the texts were transformed into numerical representations using the RoBERTa tokenizer. The information obtained from RoBERTa, along with the extracted results from Concept-LDA, were fed into the BiGRU neural network, which utilizes them for modeling temporal dependencies and better understanding of texts. Finally, an attention layer weighted important sections of the text based on their significance in analysis. Lastly, tweets were geotagged on a world map based on their publishing geography[6,7].

✉To whom correspondence may be addressed. E-mail: rahmadian.vahid@ut.ac.ir; vahid.rahmani1392@gmail.com

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

©2024 Asian Pacific Journal of Tropical Medicine Produced by Wolters Kluwer-Medknow.

How to cite this article: Jahanbin K, Jokar M, Rahmadian V. Using X Social Networks (formerly Twitter) and web news mining to predict the measles outbreak. Asian Pac J Trop Med 2024; 17(4): 188-190.

Article history: Received 20 March 2024
Accepted 23 April 2024

Revision 4 April 2024
Available online 29 April 2024

For 356 days, from 10:00 am on 17 August 2022 to 9:00 pm on 8 August 2023, tweets concerning measles were investigated on the X Social Networks.

The collected database contained 157 334 195 tweets from 115 702 540 users. Additionally, there are 2 345 018 525 users who retweeted or liked these posts and 3 246 682 910 times these posts have been viewed by users. The main hashtags used were #measles and #measles_disease.

Figure 1 illustrates the results of data mining, depicting 157 334 195 tweets from 115 702 540 users related to measles disease over a 356-day period from August 17, 2022 to August 8, 2023. Most tweets about measles are respectively in India, Pakistan, Iraq, Turkey (most in Asia, about 68 912 378 tweets accounting for 43.8%), Egypt, Sudan, Congo, South Africa, Guinea, and Morocco (most in Africa, about 36 816 202 tweets accounting for 23.4%), Ecuador, Argentina, and Chile with 22 970 793 tweets (most in South America, accounting for 14.6%), U.K, Ireland, Croatia, Italy, France, and Germany with 15 890 736 tweets (10.1%) (most in Europe), Australia with 3 304 019 (2.1%), North America like Canada, U.S.A, Mexcio, and Alaksa with 3 776 020 tweets (2.4%), and other countries with 5 664 031 tweets (3.6%) (Figure 1). This aligns with the information provided by the WHO[1]. According to the WHO, there has been a significant rise in measles cases worldwide, with a 79% increase compared to the previous year. For example, the WHO European Region reported a 30-fold increase in measles cases in 2023[1]. Furthermore, the CDC stated that the number of measles cases in the United States has already exceeded the total reported for the entire year of 2023, within just the first three months of 2024.

Infodemiology, a burgeoning field of research, delves into

scanning the Internet for user-contributed health-related content to bolster public health. This discipline involves scrutinizing online data to monitor public health issues, recognize trends, and offer insights into diverse health matters. In recent years, research has surged in the realm of extracting valuable insights from social media platforms such as X Social Networks, Google, and news websites within the domain of infodemiology. Researchers in this field analyze online content to discern sentiment, identify trends, categorize data, cluster related information, and track trends. This analytical approach plays a pivotal role in understanding how people seek health-related information online, impacting public health practices and policies. The integration of infodemiology techniques with social media analysis offers a powerful tool for researchers to gain real-time insights into public health trends and behaviors, ultimately contributing to informed decision-making in public health interventions and policies[2,4].

Recent research has demonstrated the effectiveness of using X Social Networks data for early detection and prediction of communicable and non-communicable disease outbreaks. For instance, X Social Networks mining has been used to detect influenza epidemics, predict the spread of infectious diseases such as COVID-19, Mpox, and swine flu, and develop surveillance systems for influenza, heart disease, and cancer[2,8,9]. The studies highlighted affirm that analyzing tweets is in line with recommendations from national and international healthcare bodies. X Social Networks data mining offers crucial insights into the geographic spread of cases, enabling the tracking and prediction of morbidity and mortality rates during outbreaks. Moreover, X Social Networks data analysis supports risk communication and community involvement by expediting

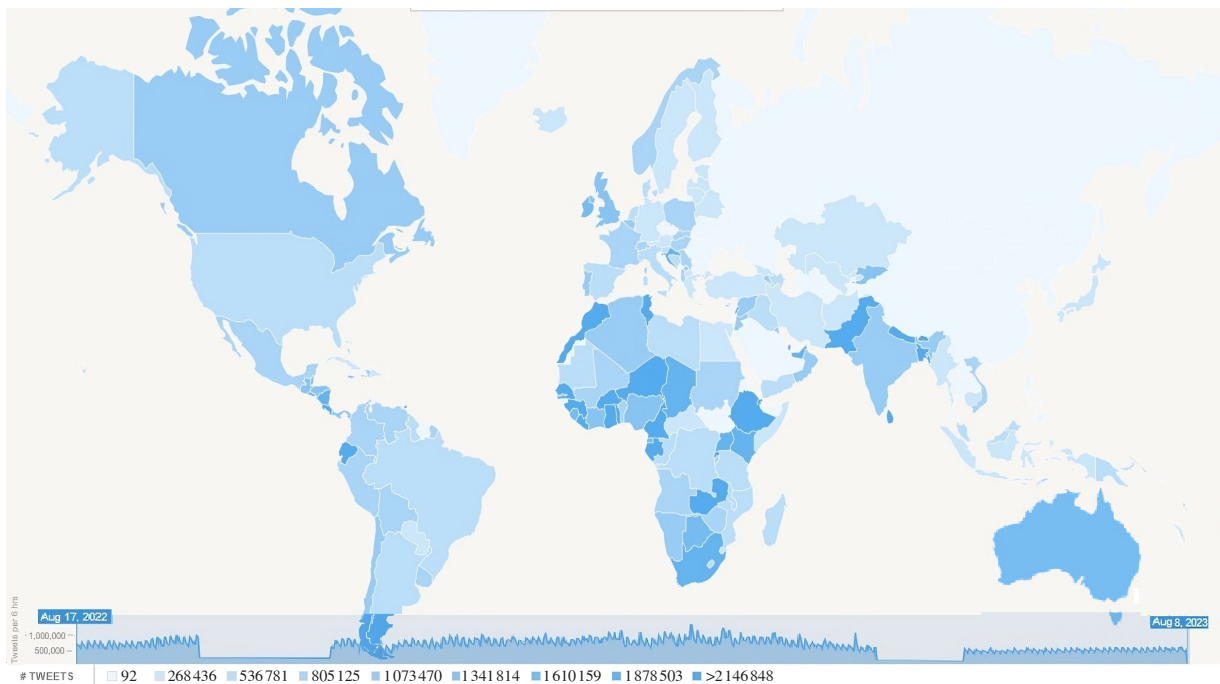


Figure 1. The geographical distribution of tweets about measles from 10:00 am on August 17, 2022 to 9:00 pm on August 8, 2023.

the evaluation of treatment options, telemedicine applications, and the monitoring of information within affected regions during epidemics[10].

Utilizing X Social Networks and web news mining techniques, this study proposes a methodology to predict measles outbreaks based on the success of similar methodologies applied to other infectious diseases, such as COVID-19 and monkeypox[2,5]. By integrating geospatial mapping and word cloud visualization techniques, we aim to identify hotspots of measles activity and uncover key themes driving disease transmission dynamics. Ultimately, we hope to facilitate more proactive and targeted public health interventions for measles control and prevention through this interdisciplinary research endeavor including improving knowledge on the measles vaccine and its cold chain management[11].

In conclusion, social media data have transformed infodemiology, providing researchers with valuable insights into human-related events. As a result of analyzing statistics provided by social networks, such as comments, photos, and videos related to social-trending diseases, it becomes possible to determine measles morbidity rates and inform health policymakers about the need for targeted educational and prevention programs in high-risk areas. Ultimately, these efforts have the potential to significantly decrease the incidence and mortality rates associated with measles in communities.

Conflict of interest statement

The authors declare no competing interests.

Funding

The authors received no extramural funding for the study.

Authors' contributions

KJ and VR conceptualised and planned the research. The task of searching and screening the literature fell to MJ. The task of gathering data fell to KJ. VR took part in the study of the data. The data interpretation was assisted by VR and MJ. The manuscript was written by MJ, KJ, and VR, who also made significant revisions. The completed work was read and approved by all writers.

References

- [1] World Health Organization. *Measles—number of reported cases*. 2024. [Online]. Available from: <https://www.who.int/data/gho/data/indicators/indicator-details/GHO/measles---number-of-reported-cases>. [Accessed on 21 March 2024].
- [2] Jahanbin K, Jokar M, Rahmanian V. Using twitter and web news mining to predict the monkeypox outbreak. *Asian Pac J Trop Med* 2022; **15**(5): 236.
- [3] Aslan Tuncay S, Akkoc G, Yilmaz S, Parlak B, Canizci Erdemli P, Dizi Isik A, et al. Epidemic management in a measles outbreak in 2023, the return of vaccine-preventable diseases: A single center, retrospective observational study. *Asian Pac J Trop Med* 2024; **17**(3): 110-118.
- [4] Jokar M, Rahmanian V, Jahanbin K. The Monkeypox outbreak reflecting rising search trend and concern in non-endemic countries: A Google trend analysis. *Disaster Med Public Health Prep* 2022; **17**: 1-3.
- [5] Jahanbin K, Rahmanian V. Using twitter and web news mining to predict COVID-19 outbreak. *Asian Pac J Trop Med* 2020; **13**(8): 378-380.
- [6] Jahanbin K, Chahooki MAZ. Aspect-based sentiment analysis of Twitter influencers to predict the trend of cryptocurrencies based on hybrid deep transfer learning models. *IEEE Access* 2023; **11**: 121656-121670.
- [7] Ekinci E, Ihan Omurca S. Concept-LDA: Incorporating BabelFy into LDA for aspect extraction. *J Inf Sci* 2020; **46**(3): 406-418.
- [8] Lee K, Agrawal A, Choudhary A. Real-time disease surveillance using twitter data: Demonstration on flu and cancer. In: *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining*. New York, NY: Association for Computing Machinery; 2013, p. 1474-1477.
- [9] Shyamala Devi N, Sharmila K, Grace Hannah J. Unmasking of heart disease symptoms using the COVID-19 vaccine dataset in Twitter. In: *Futuristic e-governance security with deep learning applications*. Hershey, PA: IGI Global; 2024, p. 191-198.
- [10] Sleigh J, Amann J, Schneider M, Vayena E. Qualitative analysis of visual risk communication on Twitter during the Covid-19 pandemic. *BMC Public Health* 2021; **21**(1): 810.
- [11] Gelaw A, Belyhun Y, Wondimeneh Y, Kokeb M, Dagne M, Amare A, et al. Knowledge and associated factors of healthcare workers on measles vaccine and cold chain management at health institutions in Gondar, Ethiopia. *Asian Pac J Trop Med* 2023; **16**(1): 26-32.

Publisher's note

The Publisher of the *Journal* remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.