LETTER TO THE EDITOR

Public perspectives of monkeypox in Twitter: A social media analysis using machine learning

Ramadan Abdelmoez Farahat
Faculty of Medicine, Kafrelsheikh University, Kafrelsheikh, 33511, Egypt

Mohammed Abdelwahab Yassin
Faculty of Medicine, South Valley University, Qena, Egypt

Jaffar A. Al-Tawfiq
Infectious Disease Unit, Specialty Internal Medicine, Johns Hopkins Aramco Healthcare, Dhahran, Saudi Arabia, Division of Infectious Diseases, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA and Division of Infectious Diseases, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Cosmin A. Bejan
Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

Basel Abdelazeem
Department of Internal Medicine, McLaren Health Care, Flint, 48532, Michigan, USA and Department of Internal Medicine, Michigan State University, East Lansing, 48823, Michigan, USA

Original Submission: 2 November 2022; Revised Submission: 15 November 2022; Accepted: 15 November 2022
Article published online: 21 November 2022

Corresponding author: Faculty of Medicine, Kafrelsheikh University, Kafrelsheikh, 33511, Egypt.
E-mail: ramadan.med_2587@med.kfs.edu.eg

Dear editor,

Monkeypox virus (MPXV) has infected more than 64,703 confirmed cases across 106 countries globally, as of October 25, 2022 [1]. Due to this concerning situation, the World Health Organization (WHO) declared the MPXV multi-country outbreak as a public health emergency of international concern (PHEIC) on July 23rd, 2022. Monkeypox disease (MPXV) is a viral zoonotic disease caused by the monkeypox virus (MPXV), which is transmitted through either animal or human close contact with infected skin, lesions, or objects (e.g., beddings, clothing, etc.), bodily blood, fluids, and respiratory droplets.

Recently, it has been reported that MPXV can be transmitted among the community of gay, bisexuals, and men who have sex with men (MSM), and this is considered a distinctive feature of the recent surge; however, further studies are needed [2]. MPXV common clinical symptoms include fever, headache, backache, myodyinia, restlessness, pathognomonic lymphadenopathy, and skin rash [2].

Given the recent spread of MPXV, associated digital information and opinions have also spread on different social media platforms, including Twitter. Determining the public trends and views about MPXV is fundamental for governments, policymakers, healthcare providers, and researchers to use the available resources to control and mitigate the burden of the recent outbreak in an efficient and timely manner [3]. Here, we performed two main analyses: 1) sentiment analysis and 2) topic modeling of social media data related to MPXV. The social media dataset was extracted through a collective search in English on Twitter Academic Application Programming Interface (API) using Machine Learning (ML) of the Python library Twint between May 22nd to August 5th, 2022, with the queries “monkeypox”, “Monkeypox cases” and “Monkeypox virus”. Sentiment analysis was used to determine the public perspective on MPXV and disease. Our approach used TextBlob, a python library that leverages Natural Language Processing to analyze the corpus contents, such as statements, comments, and tweets, to determine the overall sentiment [4]. Topic modeling was performed to assess and determine the main topics of MPXV-associated information among the general public. For this, we employed Latent Dirichlet Allocation (LDA) [5], an unsupervised learning model to aggregate words into topics. We also extracted the coherence of the results by calculating the coherence score, and U-Mass; tweets were split into a list of words (tokenization of tweets). Further, we used the Gensim Model to detect the top 30 bigrams and unigrams in our data. Lastly, for each LDA topic, we extracted keywords from the National Research Council (NRC) emotion lexicon to find emotions regarding each topic word.

Our search collected a total of 8532 tweets. From these tweets, we selected 5812 tweets for screening after discard duplicates. As a result, we obtained 5349 tweets after removing the tweets containing non-American standard code for information (ASCII) characters or only links or emails (empty tweets). Our sentiment analysis revealed 48% of tweets as neutral, 37% of tweets as positive, and 15% as negative. This analysis was also used to reveal the distribution of tweets’ polarity. It showed that the density of polarity values is highly concentrated towards zero, which indicates a more neutral
FIG. 1. (A) The distribution of tweets polarity and shows that the density of polarity values is highly concentrated towards zero, indicating a more neutral attitude. (B) Subjectivity of each tweet and shows increases towards the right (positive emotions) or left directions (negative emotions). (C) The top 30 most relevant words for topic one of the Latent Dirichlet Allocation (LDA) model which were (‘go’, ‘time’, ‘pandemic’, ‘cause’, ‘last’, ‘medium’, ‘use’, ‘then’, ‘become’, ‘sure’, ‘show’, ‘body’, ‘news’, ‘announce’, ‘real’, ‘raise’, ‘science’, ‘around’, ‘rash’, ‘research’, ‘blame’, ‘plan’, ‘past’).
TABLE 1. The distribution of the 12 topics with the top 30 most relevant words. We manually named the topic labels

| Topic label | Topic number | Keywords | Distribution (%) |
|-------------|--------------|----------|-----------------|
| The possibility of monkeypox to be pandemic | 1 | ‘go’, ‘time’, ‘pandemic’, ‘cause’, ‘last’, ‘medium’, ‘use’, ‘then’, ‘become’, ‘sure’, ‘show’, ‘body’, ‘news’, ‘announce’, ‘real’, ‘raise’, ‘science’, ‘around’, ‘risk’, ‘research’, ‘blame’, ‘plan’, ‘past’, ‘ahead’, ‘turn’, ‘break’, ‘tired’, ‘politician’, ‘supply’, ‘surpass’ | 10.405306 |
| Monkeypox origin | 2 | outbreak, woman, first, national, yet, re, likely, lab, current, too, vital, back, kind, account, play, decide, warn, th, amp, worried, one, run, maybe, instead, absolutely, fund, hold, follow, out, trend | 6.510821 |
| Gay community | 3 | monkeypox, get, vaccine, gay, people, say, emergency, virus, see, make, disease, new, still, still, global, catch, stay, country, call, week, other, community, again, risk, prevent, infectious, start, smallpox, situation, already | 4.834382 |
| Spread | 4 | so, spread, just, know, contact, child, thing, think, want, concern, also, come, contract, even, test, government, life | 2.940016 |
| Preventive measures | 5 | state, tell, stop, year, old, such, transmit, mask, as, touch, fact, different, article, bring, direct, about, lockdown, short, orgy, folk, little, flu, worry, mention, clothing, language, scab, video, breakdown, view, wear | 6.592440 |
| Declaration of the outbreak | 6 | declare, take, talk, enough, soon, ever, top, place, job, imagine, remember, monkey, seriously, listen, 1.863277 | |
| Work | 7 | now, need, breaking, here, month, right, work, shit, together, sexual, try, wait, transmission, add, well, next, fear, home, desperate, hit, hard, pay, worker, sick, scary, lose, lie, move, fast, singe | 4.040193 |
| The need for monkeypox vaccinations | 8 | more, vaccination, information, many, way, good, response, give, infection, long, tweet, literally, point, slow, discuss, datum, globally, expect, non, sexually, transmit, lase, immune, system, stigmatize, team, evidence, produce, region, check, mutate, advise | |
| Confirmed cases | 9 | case, health, confirm, public, look, patient, official, important, remain, step, very, rise, additional, city, 10.403506 | |
| Monkeypox surface amid the COVID-19 pandemic | 10 | covid, recently, most, surface, much, bad, seem, happen, learn, link, actually, include, wonder, like, 4.834382 | |
| Sexually-transmitted | 11 | crisis, isolation, repeat, able, scientist, early, case, provide, speak, common, gender, set, animal, available, game, irresponsible | 35.620724 |
| Beliefs about monkeypox | 12 | only, today, question, live, name, believe, great, idea, contain, recent, epidemic, large, almost, act, suppose, definition, ugly, history, exactly, vax, sleep, answer, right, racist, rename, attack, at, all, interview, minimze, submit | 13.729658 |

| Topic label | Topic number | Keywords | Distribution (%) |
|-------------|--------------|----------|-----------------|
| NMNI | 1 | ‘go’, ‘time’, ‘pandemic’, ‘cause’, ‘last’, ‘medium’, ‘use’, ‘then’, ‘become’, ‘sure’, ‘show’, ‘body’, ‘news’, ‘announce’, ‘real’, ‘raise’, ‘science’, ‘around’, ‘risk’, ‘research’, ‘blame’, ‘plan’, ‘past’, ‘ahead’, ‘turn’, ‘break’, ‘tired’, ‘politician’, ‘supply’, ‘surpass’ | 10.405306 |
| Monkeypox origin | 2 | outbreak, woman, first, national, yet, re, likely, lab, current, too, vital, back, kind, account, play, decide, warn, th, amp, worried, one, run, maybe, instead, absolutely, fund, hold, follow, out, trend | 6.510821 |
| Gay community | 3 | monkeypox, get, vaccine, gay, people, say, emergency, virus, see, make, disease, new, still, still, global, catch, stay, country, call, week, other, community, again, risk, prevent, infectious, start, smallpox, situation, already | 4.834382 |
| Spread | 4 | so, spread, just, know, contact, child, thing, think, want, concern, also, come, contract, even, test, government, life | 2.940016 |
| Preventive measures | 5 | state, tell, stop, year, old, such, transmit, mask, as, touch, fact, different, article, bring, direct, about, lockdown, short, orgy, folk, little, flu, worry, mention, clothing, language, scab, video, breakdown, view, wear | 6.592440 |
| Declaration of the outbreak | 6 | declare, take, talk, enough, soon, ever, top, place, job, imagine, remember, monkey, seriously, listen, 1.863277 | |
| Work | 7 | now, need, breaking, here, month, right, work, shit, together, sexual, try, wait, transmission, add, well, next, fear, home, desperate, hit, hard, pay, worker, sick, scary, lose, lie, move, fast, singe | 4.040193 |
| The need for monkeypox vaccinations | 8 | more, vaccination, information, many, way, good, response, give, infection, long, tweet, literally, point, slow, discuss, datum, globally, expect, non, sexually, transmit, lase, immune, system, stigmatize, team, evidence, produce, region, check, mutate, advise | |
| Confirmed cases | 9 | case, health, confirm, public, look, patient, official, important, remain, step, very, rise, additional, city, 10.403506 | |
| Monkeypox surface amid the COVID-19 pandemic | 10 | covid, recently, most, surface, much, bad, seem, happen, learn, link, actually, include, wonder, like, 4.834382 | |
| Sexually-transmitted | 11 | crisis, isolation, repeat, able, scientist, early, case, provide, speak, common, gender, set, animal, available, game, irresponsible | 35.620724 |
| Beliefs about monkeypox | 12 | only, today, question, live, name, believe, great, idea, contain, recent, epidemic, large, almost, act, suppose, definition, ugly, history, exactly, vax, sleep, answer, right, racist, rename, attack, at, all, interview, minimze, submit | 13.729658 |

The analysis also shows the top 30 bigrams and unigrams. The top 5 bigrams were “wear mask”, followed by “test positive”, “monkey pox”, “viral infection”, “jst amsenia” with a frequency of 10.4%, 6.5%, 5.7%, 5.2%, and 4.4%, respectively. On the contrary, the top 5 unigrams were “monkeypox”, followed by “case”, “get”, “vaccine”, and “confirm” with a frequency of 6.5%, followed by 3%, 1%, 0.98%, and 0.9%, respectively.

The bigrams cloud showed that the most frequent words are those related to outbreaks as “Monkey Pox”, “wear mask”, “Test Positive”, “Viral Infection” or those related to spread like “Sexually transmit” and “Physical contact” or related to Vaccine “Vax signup” and “side effect”.

From both unigrams and bigrams, the word clouds showed that people are talking mostly about the outbreaks and their method of spread which is likely due to physical and sexual contact.

The LDA analysis extracted 12 topics. The coherence score of the included 12 topics was about 0.4583888 with a U mass of -13.729658 because it is the most appropriate number of topics based on the data itself. Table 1 shows the distribution of the 12 topics with the top 30 words. Further, the top 30 most relevant words for topic one were shown in (Fig. 1 B).

Lastly, the distribution of selected emotion keywords from the NRC lexicon showed that positive and negative emotions – fear, anger, anticip, trust, surprise, sadness, disgust, and joy – were the most common in the 12 topics (Fig. 1 D).

Negative sentiment, fear, and sadness were the most common between topics three and four. Topic three had the emotions of negative sentiment, fear, and sadness represented 22.7%, 22.7%, and 13.6%, respectively, while topic four had negative emotions, sadness followed by fear, represented 23.5%, 17.6 %, and 11.7%, respectively. On contrary, topic nine had positive emotion followed by trust, representing 31.8% and 18%, respectively, as the most frequent.

To conclude, our sentiment analysis leveraging the NRC emotion lexicon, showed that keywords were mostly negative or expressed fear. However, TextBlob found Twitter users with a positive attitude are more common (37%) compared with those with a negative attitude (15%), while those with a neutral attitude are the main compartment (48%) which may

© 2022 The Authors. Published by Elsevier Ltd. NMNI, 49-50, 101053
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
indicate that people were not terrified about MPX. This contradiction may be due to the difference in processes of NRC, which is not a machine learning model, unlike TextBlob. Further, our sentiment analysis is limited in capturing how polarity and emotions change over time, which may reveal possible trend scenarios of transitions from negative and fearful emotions to more positive ones, for example. Our analysis also revealed that people were discussing mostly MPXV transmission, such as sexually transmitted, physical contact, the gay community, or related to the vaccine.

**Ethical approval**

Not applicable.

**Sources of funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Author contribution**

RAF: the conception and design of the study, acquisition of information, drafted the article, and final approval of the version to be submitted. MAY: the analysis of the study. RAF, MAY, JAA, CAB, and BA: interpretation of data, and revising it critically for important intellectual content. All the authors gave final approval of the version to be submitted.

**Declaration of competing interest**

All authors declare no conflict of interest, except JAA is an Associate Editor in NMNI and has no conflict of interest as well.

**Acknowledgements**

We thank Tareq AL- Ahdal for his role in providing Twitter API and data collection.

**Availability of data and materials**

Data will be provided upon request from Dr. Mohammed Abdelwahab Yassin (E-mail: muhammad.abdelwahab@med.svu.edu.eg).

**References**

[1] Kraemer MUG, Tegally H, Pigott DM, Dasgupta A, Sheldon J, Wilkinson E, et al. Tracking the 2022 monkeypox outbreak with epidemiological data in real-time. Lancet Infect Dis 2022;22:941–2. https://doi.org/10.1016/S1473-3099(22)00359-0.

[2] Hraib M, Jouni S, Albitar MM, Alaidi S, Alshehabi Z. The outbreak of monkeypox 2022: an overview. Ann Med Surg 2022;79:104069. https://doi.org/10.1016/j.amsu.2022.104069.

[3] Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020;22:e20550. https://doi.org/10.2196/20550.

[4] Sv P, Tandon J, Vikas, Hinduja H. Indian citizen’s perspective about side effects of COVID-19 vaccine – a machine learning study. Diabetes Metab Syndr Clin Res Rev 2021;15:102172. https://doi.org/10.1016/j.dsx.2021.06.009.

[5] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993–1022.