IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204116 |

Navigating Public Peceptrons of Monkeypox: A Sentiment Analysis of Twitter Discourse

J.RAMYA, DINATAKURTHY AVINASH SWAMY, GAMIDI JAGADEESH, YEDEM NITHISH KUMAR REDDY,

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT: The work presented in this paper aims to address these research challenges. First, it presents an open-access dataset of 556,427 Tweets about monkeypox that were posted on Twitter since the first detected case of this outbreak. Second, the paper reports the results of a comprehensive content analysis of the Tweets of this dataset. This analysis presents several novel findings such as − English has been the most used language (out of all the 34 languages supported by Twitter) to post Tweets about monkeypox, about 40,000 Tweets related to monkeypox were posted on the day WHO declared monkeypox as a GPHE, a total of 5470 distinct hashtags have been used on Twitter about this outbreak out of which #monkeypox is the most used hashtag, and Twitter for iPhone has been the leading source of Tweets about the outbreak. The sentiment analysis of the Tweets was also performed, and the results show that despite a lot of discussions, debate, opinions, information, and misinformation on Twitter on various topics in this regard, such as monkeypox and the LGBTQI+ community, monkeypox and COVID-19, vaccines for monkeypox, etc., "neutral" sentiment was present in most of the Tweets. It was followed by "negative" and "positive" sentiments, respectively. Finally, to support research and development in this field, the paper presents a list of 50 open research questions related to the outbreak in the areas of Big Data, Data Mining, Natural Language Processing, and Machine Learning that may be investigated based on this dataset. Future work on this research project would involve updating the dataset with more recent Tweets on a routine basis to ensure that the scientific community has access to the most recent data in this regard.

KEYWORDS: monkeypox; twitter; dataset; tweets; social media; big data

I.INTRODUCTION

Monkeypox, caused by the monkeypox virus, which belongs to the Poxviridae family, Chordopoxvirinae subfamily, and Orthopoxvirus genus, is a re-emerging zoonotic disease. The monkeypox virus was initially discovered in monkeys in 1958, and the first case of human monkeypox was detected in the Democratic Republic of the Congo (DRC) in a nine-month-old boy in 1970. The monkeypox virus is closely related to the variola virus (smallpox virus) and results in a smallpox-like disease. The incubation period of monkeypox is 5–21 days, and common symptoms include fever (between 38.5 °C and 40.5 °C), headache, and myalgia. A distinguishing feature of the monkeypox infection is the presence of swelling at the maxillary, cervical or inguinal lymph nodes (lymphadenopathy). A recent study found that during the ongoing outbreak of monkeypox, inguinal lymphadenopathy was more common than cervical and axillary lymphadenopathy. In individuals infected with the monkeypox virus, rashes appear following the onset of fever, beginning on the face, tongue, and oral cavity before spreading across the body. In the later stages of the infection, lesions in the oral cavity may make it challenging for the patients to eat and drink. However, during the ongoing outbreak, multiple atypical clinical observations have been reported as compared to the prior outbreaks. The severity of the infection is usually determined by the lesion count, as there is a direct correlation between high lesion counts and severe health-related complications [5].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204116 |

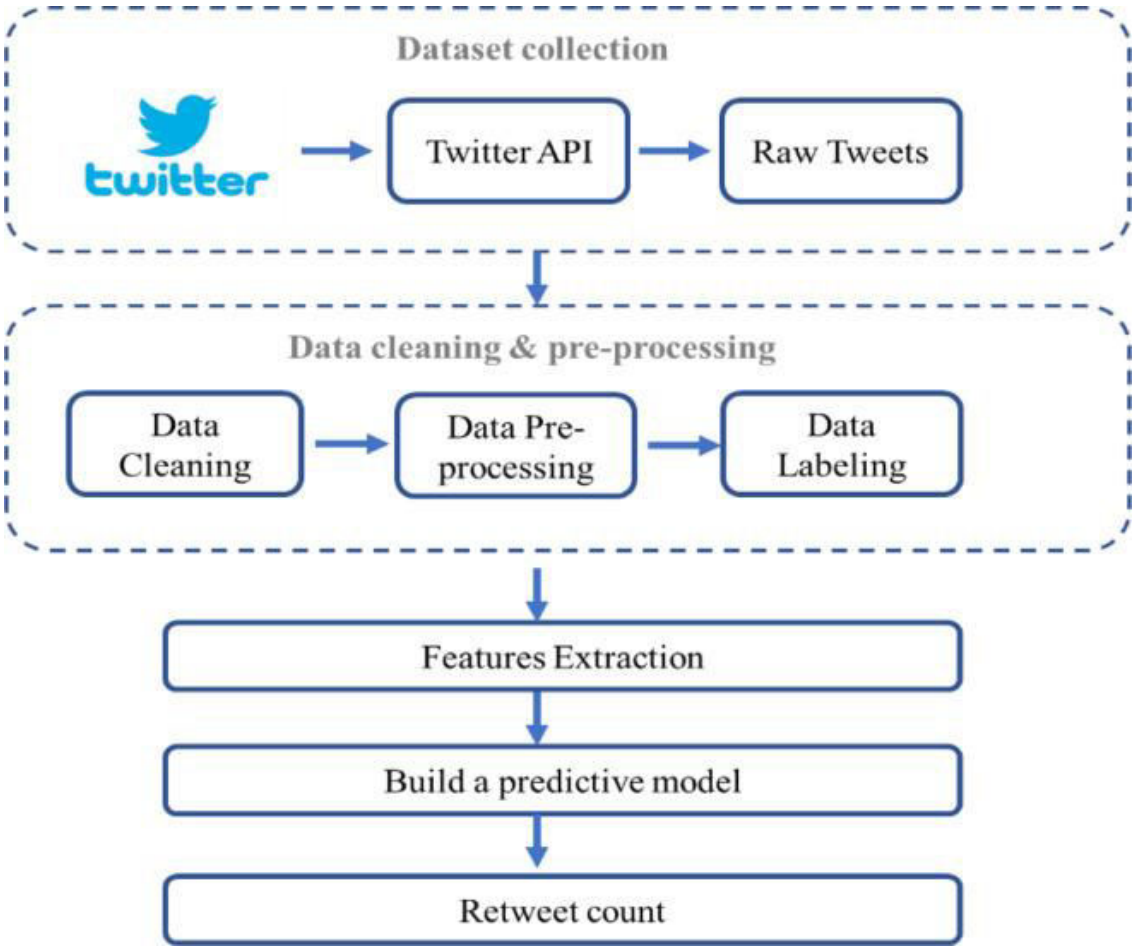

Fig 1: Predicting the popularity of tweets by analyzing

The mining of Tweets to develop datasets on recent issues, global challenges, pandemics, virus outbreaks, emerging technologies, and trending matters has been of significant interest to the scientific community in the recent past, as such datasets serve as a rich data resource for the investigation of different research questions. Furthermore, the virus outbreaks of the past, such as COVID-19, Ebola, Zika virus, and flu, just to name a few, were associated with various works related to the analysis of the multimodal components of Tweets to infer the different characteristics of conversations on Twitter related to these respective outbreaks. The ongoing outbreak of the monkeypox virus, declared a Global Public Health Emergency (GPHE) by the World Health Organization (WHO), has resulted in a surge of conversations about this outbreak on Twitter, which is resulting in the generation of tremendous amounts of Big Data. There has been no prior work in this field thus far that has focused on mining such conversations to develop a Twitter dataset. Furthermore, no prior work has focused on performing a comprehensive analysis of Tweets about this ongoing outbreak. To address these challenges, this work makes three scientific contributions to this field. First, it presents an open-access dataset of 556,427 Tweets about monkeypox that have been posted on Twitter since the first detected case of this outbreak. A comparative study is also presented that compares this dataset with 36 prior works in this field that focused on the development of Twitter datasets to further uphold the novelty, relevance, and usefulness of this dataset. Second, the paper reports the results of a comprehensive analysis of the Tweets of this dataset. This analysis presents several novel findings; for instance, out of all the 34 languages supported by Twitter, English has been the most used language to post Tweets about monkeypox, about 40,000 Tweets related to monkeypox were posted on the day WHO declared monkeypox as a GPHE, a total of 5470 distinct hashtags have been used on Twitter about this outbreak out of which #monkeypox is the most used hashtag, and Twitter for iPhone has been the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204116 |

leading source of Tweets about the outbreak. The sentiment analysis of the Tweets was also performed, and the results show that despite a lot of discussions, debate, opinions, information, and misinformation, on Twitter on various topics in this regard, such as monkeypox and the LGBTQI+ community, monkeypox and COVID-19, vaccines for monkeypox, etc., "neutral" sentiment was present in most of the Tweets. It was followed by "negative" and "positive" sentiments, respectively. Finally, to support research and development in this field, the paper presents a list of 50 open research questions related to the outbreak in the areas of Big Data, Data Mining, Natural Language Processing, and Machine Learning that may be investigated based on this dataset.

II.RELATED WORK

The mining of social media conversations, for instance, Tweets, to develop datasets has been of significant interest to the scientific community in the fields of Big Data, Data Mining, and Natural Language Processing in the last few years, as such datasets serve as a data resource for a wide range of applications related to studying the associated conversation paradigms as well as for investigating the patterns of the underlying online information-seeking and sharing behavior on Twitter. In this section, a review of such works is presented. The use cases supported by a couple of recent Twitter datasets are also outlined as examples to discuss the applicability and usefulness of Twitter datasets for the investigation of different research questions.

Mining social media conversations, for instance, Tweets, to develop datasets has been of significant interest to the scientific community in the last few years, as can be seen from several recent works in this field. Such Twitter datasets serve as a data resource for a wide range of applications and use-case scenarios related to studying the associated conversation paradigms as well as for investigating the patterns of the underlying information-seeking and sharing behavior on Twitter. Some of the recent virus outbreaks, such as COVID-19, Ebola, Zika virus, and flu, were followed by the scientific community developing Twitter datasets, performing a comprehensive analysis of the multimodal components of the Tweets (such as hashtags, language, retweets, studying the source of the Tweet, etc.), and analyzing the sentiments of these Tweets. The recent outbreak of monkeypox has also led to an increase in research and development in this field in the last few weeks). However, none of these prior works focused on mining Tweets about the 2022 monkeypox outbreak to develop a dataset. Neither did any of these prior works focus on performing a comprehensive analysis of the Tweets about this outbreak. Furthermore, there has been no work conducted in this field thus far that has focused on outlining open research questions or research directions to advance knowledge, innovation, and discovery in this field. This paper aims to address these challenges.

The rising cases of monkeypox and the associated recommendations, initiatives, and measures by various countries have led to the public engaging in conversations for information seeking and sharing related to monkeypox. The Internet of Everything lifestyle of today's living is centered around people engaging in online conversations via the internet, specifically social media platforms, and spending a lot more time on the internet than ever before. As a result, there has been a tremendous increase in the use of social media platforms in the recent past. Conversations on social media include a wide range of topics, such as recent issues, global challenges, pandemics, emerging technologies, news, current events, politics, family, relationships, trending topics, and career opportunities. Twitter, one such social media platform, is used by people of almost all age groups from different parts of the world.

III,METHODS

It allows the user to tailor search results to specific date ranges, people, and more. Specifically, the Advanced Search feature of the Twitter API allows several inputs to be provided, which include specifications to search for Tweets containing all specified keywords in any position, Tweets containing an exact phrase(s), Tweets containing any of the specified keywords, Tweets excluding specific keywords, Tweets with a specific hashtag, and Tweets in a specific language. It also allows the searching of Tweets from a specific account, Tweets sent as replies to a specific account, and Tweets that mention a specific account. This makes it easier to find specific Tweets posted during specific date ranges based on the values of one or more of these inputs.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204116 |

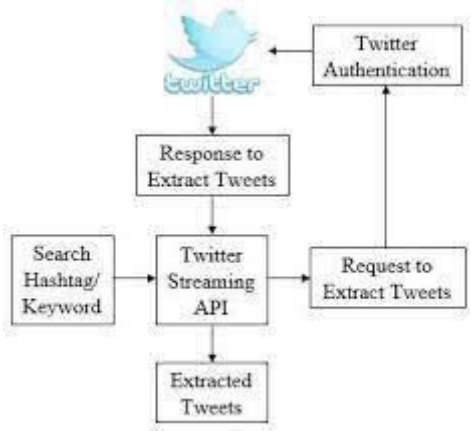

Fig 2: A Real-Time Twitter Sentiment Analysis

The results from the Advanced Search feature of the Twitter API were exported, and all the attributes from the Tweets were deleted other than the Tweet IDs to comply with the privacy policy, developer agreement, and guidelines for the content redistribution of Twitter. Thereafter, the set of Tweet IDs obtained as a result of the RapidMiner "process" was merged with the set of Tweet IDs obtained as a result of the Advanced Search feature of the Twitter API, and duplicate Tweet IDs were removed to develop this dataset. It is relevant to mention here that neither the results of Twitter API's standard search nor the results of the Advanced Search feature of the Twitter API return an exhaustive list of Tweets posted within a date range. Furthermore, Twitter users are allowed to delete a Tweet they have posted in the past. For a deleted Tweet, there will be no retrievable Tweet text and other related information upon hydration. it should have these four characteristics—Findability, Accessibility, Interoperability, and Reusability. The open-access dataset presented in this paper has a permanent and unique DOI (Section 4.1). The dataset is, therefore, findable and accessible online. The dataset files comprise only .txt files. The .txt files can be downloaded, opened, processed, and interpreted by almost all operating systems and frameworks, such as Windows, Linux, Ubuntu, Android, IOS, and so on, thereby upholding its interoperability. The dataset files present Tweet IDs. These Tweet IDs can be hydrated.

IV.RESULT ANALYSIS

A total of 139,796 Tweets were observed to be neutral, 46,586 Tweets had a negative sentiment, and 36,413 Tweets had a positive sentiment. Therefore, it can be concluded that despite a lot of discussions, debate, opinions, information, and misinformation on Twitter on various topics in this regard, such as monkeypox and the LGBTQI+ community, monkeypox and COVID-19, vaccines for monkeypox, etc., the neutral sentiment is the most common sentiment that has been associated with Tweets about the 2022 monkeypox outbreak thus far. It is worth mentioning here that Tweets posted on Twitter and their associated sentiment are quite often based on recent developments and/or events [150]. For instance, as this paper reports, about 40,000 Tweets related to monkeypox were posted on the day WHO declared monkeypox as GPHE. The monkeypox virus is spreading at a rapid rate while governments in different parts of the world and other policy-making bodies are working to develop policies to reduce the spread of the virus. In the near future, it is possible that specific policies may be recommended by certain governments and/or policy-making bodies which Twitter users of those geographic regions might be in strong disagreement or agreement with. As a result, an influx of Tweets with negative sentiments (for strong disagreement with the policies) or positive sentiments (for strong agreement with the policies) from those geographic regions could be observed on Twitter (recent examples of such Twitter activity include people in certain geographic regions using Twitter to strongly oppose the mask mandates during the early days of COVID-19 [151,152]), which might impact the overall sentiment associated with Tweets about the 2022 outbreak of monkeypox as reported in this study. To address this limitation, when the outbreak ends, this study will be repeated by including all the Tweets about monkeypox that were posted during the entire duration of the outbreak.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204116 |

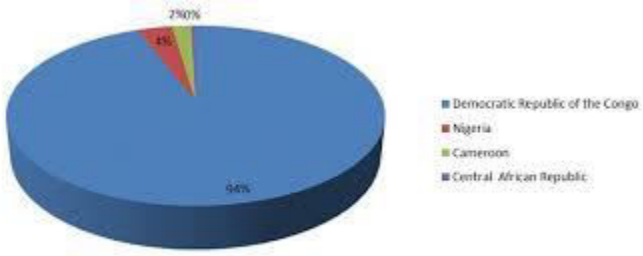

Fig 3: Result of Monkeypox cases

Based on the review of prior works in this field presented in it can be concluded that performing a comprehensive analysis of the content of the Tweets related to virus outbreaks, epidemics, and pandemics has been of significant interest to researchers in this field in the recent past. Therefore, a comprehensive content analysis of all the Tweets in this dataset was performed. The specific characteristics of the Tweets that were studied include distinct dates when the Tweets were posted, the date when the maximum number of Tweets were posted, distinct languages in which the Tweets are available, the most common language used for posting the Tweets, the total number of different hashtags present in all the Tweets, most commonly used hashtag, the percentage of Tweets posted using an iPhone (Twitter for iPhone), the percentage of Tweets posted using an Android phone (Twitter for Android), and the percentage of Tweets posted using the Twitter website (Twitter Web App). This analysis was performed using RapidMiner] using its data analysis features after hydrating the Tweet IDs.

V.CONCLUSIONS

Twitter datasets serve as a rich data resource for the investigation of different research questions for the timely advancement of knowledge, innovation, and discovery in different fields. Therefore, scientists in this field have focused on developing Twitter datasets on recent issues, global challenges, pandemics, virus outbreaks, emerging technologies, and trending matters in the last few years. In addition to the development of Twitter datasets, analysis of multimodal components of Tweets, specifically Tweets about virus outbreaks, has been of significant interest to the scientific community, as can be seen from several works that focused on analyzing different characteristics of Tweets posted about some of the recent virus outbreaks, such as COVID-19, Ebola, Zika virus, and the flu. The world is currently experiencing an outbreak of the monkeypox virus. A total of 71,096 cases have been reported so far, out of which 70,377 cases have been reported in locations that have not historically reported any monkeypox infections. The World Health Organization (WHO) has declared monkeypox to be a Global Public Health Emergency. This has resulted in a tremendous increase in different types of conversations on Twitter related to monkeypox. None of the prior works in this field have focused on mining these conversations to develop a Twitter dataset.

REFERENCES

1. McCollum, A.M.; Damon, I.K. Human Monkeypox. *Clin. Infect. Dis.* **2014**, *58*, 260–267. [Google Scholar] [CrossRef] [PubMed] [Green Version]
2. von Magnus, P.; Andersen, E.K.; Petersen, K.B.; Birch-Andersen, A. A Pox-like Disease in Cynomolgus Monkeys. *Acta Pathol. Microbiol. Scand.* **2009**, *46*, 156–176. [Google Scholar] [CrossRef]
3. Breman, J.G.; Kalisa-Ruti; Steniowski, M.V.; Zanotto, E.; Gromyko, A.I.; Arita, I. Human Monkeypox, 1970–1979. *Bull. World Health Organ.* **1980**, *58*, 165–182. [Google Scholar] [PubMed]
4. Charniga, K.; Masters, N.B.; Slayton, R.B.; Gosdin, L.; Minhaj, F.S.; Philpott, D.; Smith, D.; Gearhart, S.; Alvarado-Ramy, F.; Brown, C.; et al. Estimating the Incubation Period of Monkeypox Virus during the 2022 Multi-National Outbreak. *medRxiv* **2022**, arXiv:2022.06.22.22276713. [Google Scholar]
5. Jezek, Z.; Szczeniowski, M.; Paluku, K.M.; Mutombo, M. Human Monkeypox: Clinical Features of 282 Patients. *J. Infect. Dis.* **1987**, *156*, 293–298. [Google Scholar] [CrossRef] [PubMed]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204116 |

6. Perez Duque, M.; Ribeiro, S.; Martins, J.V.; Casaca, P.; Leite, P.P.; Tavares, M.; Mansinho, K.; Duque, L.M.; Fernandes, C.; Cordeiro, R.; et al. Ongoing Monkeypox Virus Outbreak, Portugal, 29 April to 23 May 2022. *Euro Surveill.* **2022**, *27*, 2200424. [Google Scholar] [CrossRef]
7. Antinori, A.; Mazzotta, V.; Vita, S.; Carletti, F.; Tacconi, D.; Lapini, L.E.; D'Abramo, A.; Cicalini, S.; Lapa, D.; Pittalis, S.; et al. Epidemiological, Clinical and Virological Characteristics of Four Cases of Monkeypox Support Transmission through Sexual Contact, Italy, May 2022. *Euro Surveill.* **2022**, *27*, 2200421. [Google Scholar] [CrossRef]
8. Hammerschlag, Y.; MacLeod, G.; Papadakis, G.; Adan Sanchez, A.; Druce, J.; Taiaroa, G.; Savic, I.; Mumford, J.; Roberts, J.; Caly, L.; et al. Monkeypox Infection Presenting as Genital Rash, Australia, May 2022. *Euro Surveill.* **2022**, *27*, 2200411. [Google Scholar] [CrossRef]
9. Huhn, G.D.; Bauer, A.M.; Yorita, K.; Graham, M.B.; Sejvar, J.; Likos, A.; Damon, I.K.; Reynolds, M.G.; Kuehnert, M.J. Clinical Characteristics of Human Monkeypox, and Risk Factors for Severe Disease. *Clin. Infect. Dis.* **2005**, *41*, 1742–1751. [Google Scholar] [CrossRef]
10. Adler, H.; Gould, S.; Hine, P.; Snell, L.B.; Wong, W.; Houlihan, C.F.; Osborne, J.C.; Rampling, T.; Beadsworth, M.B.; Duncan, C.J.; et al. Clinical Features and Management of Human Monkeypox: A Retrospective Observational Study in the UK. *Lancet Infect. Dis.* **2022**, *22*, 1153–1162. [Google Scholar] [CrossRef]

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details