Communication

MonkeyPox2022Tweets: The First Public Twitter Dataset on the 2022 MonkeyPox Outbreak

Nirmalya Thakur

Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221-0030, U.S.A.; thakurna@mail.uc.edu

Abstract: The world is currently facing an outbreak of the monkeypox virus and confirmed cases have been reported from 74 countries. Following a recent "emergency meeting", the World Health Organization just declared monkeypox a global health emergency. As a result, people from all over the world are using social media platforms, such as Twitter, for information seeking and sharing related to the outbreak, as well as for familiarizing themselves with the guidelines and protocols that are being recommended by various policy-making bodies to reduce the spread of the virus. This is resulting in the generation of tremendous amounts of Big Data related to such paradigms of social media behavior. Mining this Big Data and compiling it in the form of a dataset can serve a wide range of use-cases and applications such as analysis of public opinions, interests, views, perspectives, attitudes, and sentiment towards this outbreak. Therefore, this work presents MonkeyPox2022Tweets, an open-access dataset of more than 255,000 Tweets related to the 2022 monkeypox outbreak that were posted on Twitter since the first detected case of this outbreak on May 7, 2022. The dataset is compliant with the privacy policy, developer agreement, and guidelines for content redistribution of Twitter, as well as with the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management.

Keywords: Monkeypox; monkey pox; Twitter; Dataset; Tweets; Social Media; Big Data; Data Mining; Data Science

1. Introduction

Monkeypox, caused by the monkeypox virus, which belongs to the Poxviridae family, Chordopoxvirinae subfamily, and Orthopoxvirus genus [1], is a re-emerging zoonotic disease. The monkeypox virus was initially discovered in monkeys in 1958 [2], 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 [3]. The monkeypox virus is closely related to the variola virus (smallpox virus) and results in a smallpox-like disease. The symptoms usually develop within two weeks of exposure. The initial symptoms usually comprise fever, headache, muscle aches and backache, swollen lymph nodes, chills, and exhaustion. Within 1-3 days of these initial symptoms, most infected individuals develop a rash or a sore, which usually appears on the face, followed by rapid spreading to different parts of the body in a centrifugal manner [4]. There are two different clades of this virus: the Congo Basin (central African) clade and the West African clade. These two clades have estimated fatality rates of 10.6% and 3.6%, respectively [5].

The virus had been endemic in the DRC and a few African countries for a very long time, and a few cases outside these geographic regions were recorded only twice – first in 2003 [6] and then in 2018-19 [7,8]. However, at the time of writing this paper, the world is experiencing a global outbreak of the monkeypox virus, with a total of 16,836 confirmed cases in 74 countries [9]. The first case of this 2022 global monkeypox outbreak was confirmed in the United Kingdom on May 7, 2022 [10]. On May 19, 2022, the first draft genome sequence of the monkeypox virus was performed by scientists in Portugal [11]. The genomic data related to this outbreak that has been studied so far indicates that this outbreak is caused by the West African clade [12].

On May 20, 2022, the World Health Organization (WHO) called an "emergency meeting" [13] to discuss the global concerns centered around the rising cases of the monkeypox virus. Over the next few days, the WHO considered whether the outbreak should be assessed as a "potential public health emergency of international concern" or PHEIC, as was done for the COVID-19 and Ebola outbreaks in the past [14]. On June 6, 2022, the Center for Disease Control (CDC) in the United States raised its monkeypox alert to "Level 2" following the rapid increase in cases [15]. After another meeting, on July 23, 2022, the WHO declared monkeypox a global health emergency.

As per CDC, "at this time, there are no specific treatments available for monkeypox infection" [16]. However, recently, a vaccine for monkeypox has been approved by the Food and Drug Association (FDA). The vaccine, previously used for smallpox, is called Imvamune or Imvanex or the Jynneos vaccine and was developed by Bavarian Nordic, a Danish biotechnology firm [17]. At the time of writing this paper, the Jynneos vaccine is the only F.D.A.-approved vaccine against monkeypox. Tecovirimat (Tpoxx), an antiviral medication, was approved in the U.S.A., Canada, and Europe for the treatment of human smallpox, and the European Medicines Agency has just approved the same for the treatment of both smallpox and monkeypox [18]. As the cases surge, countries all over the world are taking various forms of preparations, initiatives, and measures to reduce the spread of the virus. These include a lockdown in Belgium [19], the United States ordering 500,000 does of the Jynneos vaccine [20], Canada offering vaccination to high-risk groups [21], health authorities in France and Denmark wanting a vaccine rollout to adults infected by the virus [22], Germany recommending vaccinations for high-risk groups [23], and the United Kingdom advising self-isolation for everyone infected with the virus [24].

The rising cases of monkeypox and the associated recommendations, initiatives, and measures by various countries have led to the general 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 [25]. As a result, there has been a tremendous increase in the use of social media platforms in the recent past [26]. Conversations on social media include a wide range of topics such as recent issues, global challenges, emerging technologies, news, current events, politics, family, relationships, and career opportunities [27]. Twitter, one such social media platform, is used by people of almost all age groups from all parts of the world [28,29]. At present, there are about 450 million monthly active users on Twitter [30].

Mining of 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 these recent works where relevant Tweets were mined to develop Twitter datasets on COVID-19 [31,32], 2022 war between Ukraine and Russia [33,34], European Migration Crisis [35], Inflammatory Bowel Disease [36], toxic behavior amongst adolescents [37], music [38], civil unrest [39], drug safety [40], and movies [41].

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. For instance, the Twitter dataset on music [38] has helped in the development of a context-aware music recommendation system [42], next-track music recommendations as per user personalization [43], session-based music recommendation algorithms [44], music recommendation systems based on the use of affective hashtags [45], music chart predictions [46], user curated playlists [47], sentiment analysis of music [48], listener engagement with popular songs [49], culture aware music recommendation systems [50], mining of user personalities [51], and several other applications. Similarly, the COVID-19 Twitter dataset proposed in [32] has been used for topic detection and sentiment analysis [52], language-agnostic discourse classification of tweets [53], monitoring emotions of people [54], studying the dynamics of user emotions [55], conversation analysis [56], studying the evolution of public sentiments over time [57], and several other applications.

The recent outbreak of monkeypox has also led to an increase in research and development in this field. These include - a study on the outbreaks in Europe and North America [58], an analysis of stigmatization of the LGBTQI+ community due to the outbreak [59], studying the increasing cases in England [60], research on monkeypox image data [61], investigating methods of transmission of the virus through sexual contact [62], analyzing public attitude towards monkeypox [63], predicting the burden and duration of this outbreak [64], use of the Jynneos vaccine for preexposure vaccination of people at risk for occupational exposure [65], nanopore sequencing of a monkeypox virus strain [66], studying the incubation period [67], guidelines for pregnant individuals with monkeypox virus exposure [68], attitudes of the US general public towards monkeypox [69], a rapid metagenomic method to detect emerging viral pathogens applied to human monkeypox [70], and information that dermatologists need to know regarding this outbreak [71].

However, none of these works have focused on mining Tweets about the 2022 Monkeypox outbreak. To address these limitations, this work proposes an open-access dataset of more than 255,000 Tweet IDs of the same number of Tweets about monkeypox that were posted on Twitter from May 7, 2022, to July 23, 2022. The earliest date was selected as May 7, 2022, as the first case of the 2022 monkeypox outbreak was recorded on this date. July 23, 2022 was the most recent date as per the time of data collection and writing of this paper.

The rest of the paper is organized as follows. The methodology that was followed for the development of this dataset is presented in Section 2. This section also outlines how the dataset is compliant with the privacy policy, developer agreement, and guidelines for content redistribution of Twitter, as well as with the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management. Section 3 presents the results of this work. It describes the dataset files as well as presents step-by-step instructions on how to use this dataset. This section also includes a compilation of multiple research questions related to the outbreak that may be using this dataset. The conclusion and scope for future work are presented in Section 4, which is followed by references.

2. Materials and Methods

This section is divided into three parts. Section 2.1 presents the specific steps that were followed for the development of this dataset. Section 2.2 outlines how this dataset is compliant with the privacy policy, developer agreement, and guidelines for content redistribution of Twitter. Section 2.3 states the compliance of this dataset with the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management.

2.1. Steps for Dataset Development

The dataset was developed by searching tweets that comprised the keyword(s) “monkeypox” or “monkey pox”, posted from May 7, 2022, to July 23, 2022 (the most recent date at the time of writing this paper). This search and the associated mining of tweets were performed as per Twitter API’s standard search policies [81]. There are various tools and applications available that comply with these policies and help to search tweets based on one or more keywords. The specific tool that was used for this work is RapidMiner [82]. RapidMiner was used because of its easy-to-use integrated development environment that allows the development of a range of Big Data and Data Mining-based applications using a combination of both built-in and user-defined functionalities. These built-in functionalities are available in the form of “operators” that can be customized as well as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2022 doi:10.20944/preprints202206.0172.v3

integrated for developing a working application on the RapidMiner platform, known as a “process”. The platform also allows the user to develop an “operator” from scratch and bundle the same with other built-in or user-defined “operators” to develop a “process”.

For this work, RapidMiner studio, version 9.10, was downloaded and installed on a laptop with the Microsoft Windows 10 Home operating system with Intel(R) Pentium(R) Silver N5030 CPU @ 1.10GHz, 1101 Mhz, 4 Core(s), and 4 Logical Processor(s). The specific functionality that was required for this work was searching tweets based on the matching keyword(s) within a date range. This functionality is already available in RapidMiner Studio 9.10 as a built-in “operator” called the Search Twitter “operator” [83] that works by connecting with the Twitter API and by complying with the Twitter API’s standard search policies for searching relevant tweets. Here, relevant tweets are defined as those tweets which contain the keyword(s) that are entered as an input to this “operator”. So, a “process” was developed in RapidMiner that comprised only the Search Twitter “operator,” and it was used to search tweets that contained either “monkeypox” or “monkey pox”, posted on Twitter in the date range May 7, 2022, to July 23, 2022. This process was run multiple times in this date range to collect the relevant tweets. A total of 255,363 tweets were mined by following this approach.

The result of this RapidMiner "process" comprised multiple attributes – "Row no", "Id", "Created-At", "From-User", "From-User-Id", "To-User", "To-User-Id", "Language", "Source Text", "Geo-Location-Latitude", "Geo-Location-Longitude", and "Retweet Count". These refer to the row number of the results, tweet ID of the obtained tweet, date and time when the tweet was posted, the username of the Twitter user who posted the tweet, the user ID of the Twitter user who posted the tweet, Twitter username of the user whose tweet was replied to (if the tweet was a reply) in the current tweet, Twitter user ID of the user whose tweet was replied to (if the tweet was a reply) in the current tweet, the language of the tweet, source of the tweet to determine if the tweet was posted from an Android source, Twitter website, etc., the complete text of the tweet, including embedded URLs, geo-location (latitude) of the user posting the tweet, geo-location (longitude) of the user posting the tweet, and retweet count of the tweet. To comply with the privacy policy, developer agreement, and guidelines for content redistribution of Twitter [76,77], multiple data filters were introduced in the RapidMiner “process” to remove all the attributes from the results other than the “Id” attribute. Thereafter the results from multiple runs of this “process” were exported and compiled to develop this dataset.

The screenshot of the ”process” that was developed in RapidMiner for the development of this dataset is shown in Figure 1. Table 1 outlines the functionality of the “operators” of this RapidMiner “process”.

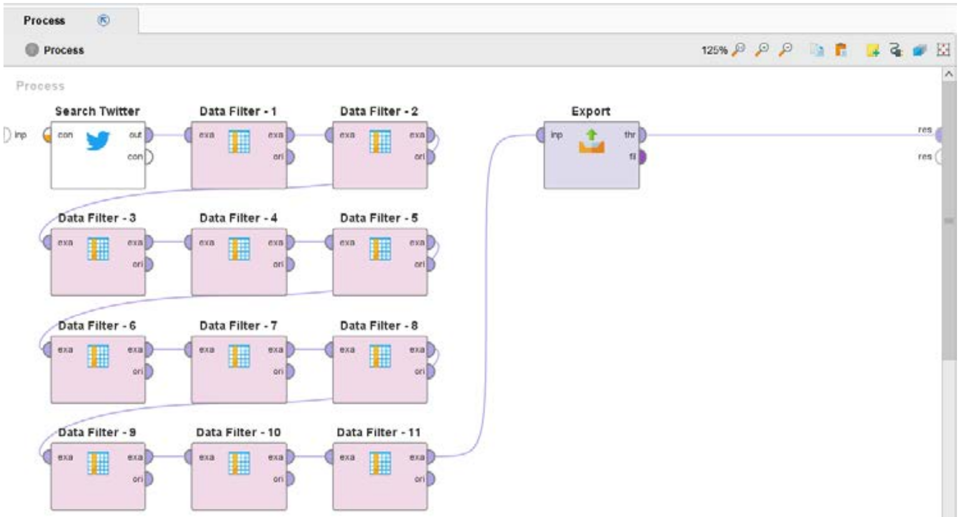

Figure 1. The RapidMiner “process” that was used for the development of this dataset

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2022 doi:10.20944/preprints202206.0172.v3

Table 1. Description of the "operators" of the RapidMiner "process" that was used for the development of this dataset.

Operator Name	Description
Search Twitter	Searches relevant tweets from Twitter by connecting with the Twitter API and by complying with the Twitter API's standard search policies
Data Filter - 1	Removes the attribute that contains the date and time when the tweet was posted
Data Filter - 2	Removes the attribute that contains the Twitter username of the user who posted the tweet
Data Filter - 3	Removes the attribute that contains the Twitter User ID of the user who posted the tweet
Data Filter - 4	Removes the attribute that contains the Twitter username of the user whose tweet was replied to (if the tweet was a reply) in the current tweet
Data Filter - 5	Removes the attribute that contains the Twitter user ID of the user whose tweet was replied to (if the tweet was a reply) in the current tweet
Data Filter - 6	Removes the attribute that contains the language of the tweet
Data Filter - 7	Removes the attribute that contains the source of the tweet, such as an Android source, Twitter website, etc.
Data Filter - 8	Removes the attribute that contains the complete text of the tweet, including embedded URLs
Data Filter - 9	Removes the attribute that contains the Geo-Location (Latitude) of the user posting the tweet
Data Filter - 10	Removes the attribute that contains Geo-Location (Longitude) of the user posting the tweet
Data Filter - 11	Removes the attribute that contains the retweet count of the tweet
Export	Exports the result as a .csv file on the local computer

It is relevant to mention here that Twitter API's standard search does not 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 (Section 3) of that Tweet ID. At the time of writing this paper, all 255,363 Tweet IDs present in this dataset, corresponded to Tweets that were not deleted.

2.2. Compliance with Twitter Policies

The privacy policy of Twitter [76] states – *"Twitter is public and Tweets are immediately viewable and searchable by anyone around the world"*. The guidelines for Twitter content redistribution [77] state – *"If you provide Twitter Content to third parties, including downloadable datasets or via an API, you may only distribute Tweet IDs, Direct Message IDs, and/or User IDs (except as described below)*. It also states - *"We also grant special permissions to academic researchers sharing Tweet IDs and User IDs for non-commercial research purposes. Academic researchers are permitted to distribute an unlimited number of Tweet IDs and/or User IDs if they are doing so on behalf of an academic institution and for the sole purpose of non-commercial research."* Therefore, it may be concluded that mining relevant tweets from Twitter to develop a dataset (comprising only Tweet IDs) to share the same is in compliance with the privacy policy, developer agreement, and content redistribution guidelines of Twitter.

2.3. Compliance with FAIR

The FAIR principles for scientific data management [84] state that a dataset should have Findability, Accessibility, Interoperability, and Reusability. The dataset is findable as it has a unique and permanent DOI. The dataset is accessible online. It is interoperable due to the use of .txt files for data representation that can be downloaded, read, and analyzed across different computer systems and applications. The dataset is re-usable as the associated tweets and related information, such as user ID, user name, retweet count, etc., for all the Tweet IDs can be obtained by the process of hydration in compliance with Twitter policies (Section 3.2) for data analysis and interpretation.

3. Results and Discussion

This section presents the results and findings of this work. It is divided into three parts. Section 3.1 presents the description of the dataset files. A step-by-step instruction on how to use this dataset is presented in Section 3.2. A list of potential research questions

 doi:10.20944/preprints202206.0172.v3

related to the fields of Big Data, Data Mining, and Natural Language Processing, that may be investigated using this dataset is presented in Section 3.3.

3.1. Description of the Dataset Files

This dataset is publicly available at https://doi.org/10.5281/zenodo.6898178. At the time of submission of this paper, in just one month, since the publication of the preprint of the paper on the preprints.org platform, the preprint has been downloaded 272 times [72], and the dataset has received a total of 2704 views (views on IEEE Dataport: 2119 [73], views on Zenodo: 229 [74], views on Kaggle: 356 [75]). The dataset consists of a total of 255,363 tweet IDs of tweets about monkeypox that were posted on Twitter from May 7, 2022, to July 23, 2022. The Tweet IDs are presented in 6 different .txt files based on the timelines of the associated tweets. Table 1 provides the details of these dataset files.

Table 1. Description of all the files present in this dataset.

Filename	No. of Tweet IDs	Date Range of the Tweet IDs
TweetIDs_Part1.txt	13926	May 7, 2022 to May 21, 2022
TweetIDs_Part2.txt	17705	May 21, 2022 to May 27, 2022
TweetIDs_Part3.txt	17585	May 27, 2022 to June 5, 2022
TweetIDs_Part4.txt	19718	June 5, 2022 to June 11, 2022
TweetIDs_Part5.txt	47718	June 12, 2022 to June 30, 2022
TweetIDs_Part6.txt	138711	July 1, 2022 to July 23, 2022

To comply with the privacy policy, developer agreement, and guidelines for content redistribution of Twitter [76,77], only the Tweet IDs associated with these 255,363 tweets are presented in this dataset. To obtain the detailed information associated with each of these tweets, such as the tweet text, user name, user ID, timestamp, retweet count, etc., these Tweet IDs need to be hydrated. There are several applications, such as the Hydrator app [78], Social Media Mining Toolkit [79], and Twarc [80], that work by complying with Twitter policies and may be used for hydrating the Tweet IDs in this dataset. A step-by-step process for using one of these applications – the Hydrator app for hydrating the files in this dataset, is presented in Section 3.2.

3.2. Instructions for Using the Dataset

The following is the step-by-step process for using the Hydrator app [78] to hydrate this dataset or, in other words, to obtain the text of the tweet, user ID, user name, retweet count, language, tweet URL, source, and other public information related to all the Tweet IDs present in this dataset. The Hydrator app works in compliance with the policies for accessing and calling the Twitter API.

1. Download and install the desktop version of the Hydrator app from https://github.com/DocNow/hydrator/releases.
2. Click on the "Link Twitter Account" button on the Hydrator app to connect the app to an active Twitter account.
3. Click on the "Add" button to upload one of the dataset files (such as TweetIDs_Part4.txt). This process adds the dataset file to the Hydrator app.
4. If the file upload is successful, the Hydrator app will show the total number of Tweet IDs present in the file. For instance, for the file - "TweetIDs_Part4.txt", the app would show the Number of Tweet IDs as 19718.
5. Provide details for the respective fields: Title, Creator, Publisher, and URL in the app, and click on "Add Dataset" to add this dataset to the app.
6. The app would automatically redirect to the "Datasets" tab. Click on the "Start" button to start hydrating the Tweet IDs. During the hydration process, the progress indicator would increase, indicating the number of Tweet IDs that have

been successfully hydrated and the number of Tweet IDs that are pending hydration.

7. After the hydration process ends, a .jsonl file would be generated by the app that the user can choose to save on the local storage.
8. The app would also display a "CSV" button in place of the "Start" button. Clicking on this "CSV" button would generate a .csv file with detailed information about the tweets, which would include the text of the tweet, user ID, user name, retweet count, language, tweet URL, source, and other public information related to the tweet.
9. Repeat steps 3-8 for hydrating all the files of this dataset.

3.3. Potential Research Questions

The recent works in the fields of Big Data, Data Mining, and Natural Language Processing, related to Twitter data analysis [42-57] and the development of Twitter datasets [31-41], as discussed in Section 1, uphold the fact that Twitter datasets serve as a rich data resource for the investigation of research questions on a wide range of topics as well as for different use case scenarios. Based on studying these recent works, the following is a list of potential research questions related to the 2022 monkeypox outbreak that may be investigated by using this dataset:

1. What is the overall emotion (positive, negative, or neutral) of the general public related to the outbreak as expressed on Twitter?
2. Which machine learning classifier (such as Random Forest, Decision Trees, Naïve Bayes, etc.) achieves the best performance accuracy for sentiment analysis of tweets related to monkeypox?
3. Are there any specific aspects or subject matters related to the outbreak (such as vaccines, treatments, and protocols to reduce the spread) that are consistently associated with a positive (or negative) emotion on Twitter?
4. What are some of the commonly used hashtags and trends in the same related to tweets about the outbreak?
5. Have there been any trending discussions on Twitter related to one or more matters (such as new protocols to reduce the spread) concerning the outbreak?
6. Has Twitter played a role in the development and spread of any conspiracy theories about monkeypox?
7. Are any political leaders or popular personalities using Twitter to spread misinformation or fake news related to monkeypox?
8. How is Twitter being used by news organizations, including regional media, local media, national media, and broadcast news agencies, in the dissemination of the latest developments related to the outbreak?
9. What were the specific characteristics of the tweets (character count, embedded URLs, date and time stamp, etc.) about monkeypox that were retweeted the most?
10. Can the tweets be analyzed to develop a machine learning classifier that would indicate the degree of credibility and accuracy of information about monkeypox expressed in these tweets from different sources?
11. What are some of the concerns or needs or complaints about the outbreak expressed by people on Twitter from different geographic regions?
12. Is there any pattern of emoji usage in the tweets about monkeypox since the beginning of the outbreak?
13. Is there any correlation between the number of tweets about monkeypox from a geographic region and the number of cases in the same region?
14. What is the best time to tweet (in terms of highest user engagement) about a new policy, measure, protocol, or news about monkeypox?
15. Can the content of the tweets be studied to investigate any potential online stigmatization, discrimination, and/or hate faced by any diversity group, such as the LGBTQI+ community?

16. Do the tweets reveal any form of panic behavior (such as panic buying of certain products as was observed during COVID-19) in regions with a high number of cases?
17. Is there any feedback that individuals infected with the virus have communicated on Twitter related to the treatment they received?
18. Can the tweets be studied to infer stress or anxiety in individuals tweeting about the virus who are experiencing one of more symptoms after getting infected?
19. What are some of the most popular websites (such as specific news outlets) that have been shared the most on Twitter for sharing and exchange of information about monkeypox?
20. Can the tweets be analyzed to develop different user personas in terms of the underlining views, opinions, and perspectives about monkeypox expressed in the tweets?

4. Conclusion and Future Scope of Work

Since the first case on May 7, 2022, the monkeypox virus has infected 16,836 people in 74 different countries, and the number of cases is increasing. Following a recent "emergency meeting", the World Health Organization just declared monkeypox as a global health emergency. As a result, several countries are implementing various forms of measures, policies, and guidelines to reduce the spread of the virus. These policies and the increase in cases on a global scale have led to an increase in conversations on social media, specifically Twitter, related to information seeking and sharing about monkeypox. These conversations are leading to the generation of tremendous amounts of Big Data. Mining Twitter conversations on specific topics or viruses or global challenges to develop datasets has been of significant interest to the scientific community in the fields of Big Data, Data Mining, Natural Language Processing, and their related areas in the last few years. While there have been several Twitter datasets developed on different topics in the past, none of those Twitter datasets are a collection of tweets related to the current outbreak of monkeypox. Furthermore, none of the works related to the 2022 monkeypox outbreak have focused on the development of any such datasets thus far. To address this challenge, this work presents MonkeyPox2022Tweets, an open-access dataset of more than 255,000 Tweet IDs of the same number of Tweets about monkeypox that were posted on Twitter from May 7, 2022, to July 23, 2022 (the most recent date at the time of writing this paper). The paper also presents several research questions that may be investigated using this dataset. Future work on this project would involve updating the dataset with more recent tweets to ensure that the scientific community has access to the recent data in this regard.

Supplementary Materials: Not applicable.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The data presented in this study are publicly available at https://doi.org/10.5281/zenodo.6898178

Conflicts of Interest: The author declares no conflict of interest

References

1. McCollum, A.M.; Damon, I.K. Human Monkeypox. *Clin. Infect. Dis.* **2014**, *58*, 260–267, doi:10.1093/cid/cit703.
2. Magnus, P. von; Andersen, E.K.; Petersen, K.B.; Birch-Andersen, A. A Pox-like Disease in Cynomolgus Monkeys. *Acta Pathol. Microbiol. Scand.* **2009**, *46*, 156–176, doi:10.1111/j.1699-0463.1959.tb00328.x.
3. Breman, J.G.; Kalisa-Ruti; Steniowski, M.V.; Zanotto, E.; Gromyko, A.I.; Arita, I. Human Monkeypox, 1970-79. *Bull. World Health Organ.* **1980**, *58*, 165–182.

4. Vaughan, A.; Aarons, E.; Astbury, J.; Brooks, T.; Chand, M.; Flegg, P.; Hardman, A.; Harper, N.; Jarvis, R.; Mawdsley, S.; et al. Human-to-Human Transmission of Monkeypox Virus, United Kingdom, October 2018. *Emerg. Infect. Dis.* **2020**, *26*, 782–785, doi:10.3201/eid2604.191164.
5. Likos, A.M.; Sammons, S.A.; Olson, V.A.; Frace, A.M.; Li, Y.; Olsen-Rasmussen, M.; Davidson, W.; Galloway, R.; Khristova, M.L.; Reynolds, M.G.; et al. A Tale of Two Clades: Monkeypox Viruses. *J. Gen. Virol.* **2005**, *86*, 2661–2672, doi:10.1099/vir.0.81215-0.
6. Centers for Disease Control and Prevention (CDC.) Multistate Outbreak of Monkeypox--Illinois, Indiana, and Wisconsin, 2003. *MMWR Morb. Mortal. Wkly. Rep.* **2003**, *52*, 537–540.
7. Public Health England Monkeypox Case Confirmed in England Available online: https://www.gov.uk/government/news/monkeypox-case-confirmed-in-england (accessed on 12 June 2022).
8. Erez, N.; Achdout, H.; Milrot, E.; Schwartz, Y.; Wiener-Well, Y.; Paran, N.; Politi, B.; Tamir, H.; Israely, T.; Weiss, S.; et al. Diagnosis of Imported Monkeypox, Israel, 2018. *Emerg. Infect. Dis.* **2019**, *25*, 980–983, doi:10.3201/eid2505.190076.
9. 2022 Monkeypox Outbreak Global Map. Available online: https://www.cdc.gov/poxvirus/monkeypox/response/2022/world-map.html (accessed on 24 July 2022).
10. Saxena, S.K.; Ansari, S.; Maurya, VK; Kumar, S.; Jain, A.; Paweska, J.T.; Tripathi, A.K.; Abdel-Moneim, A.S. Re-Emerging Human Monkeypox: A Major Public-Health Debacle. *J. Med. Virol.* **2022**, doi:10.1002/jmv.27902.
11. First Draft Genome Sequence of Monkeypox Virus Associated with the Suspected Multi-Country Outbreak, May 2022 (Confirmed Case in Portugal) Available online: https://virological.org/t/first-draft-genome-sequence-of-monkeypox-virus-associated-with-the-suspected-multi-country-outbreak-may-2022-confirmed-case-in-portugal/799 (accessed on 12 June 2022).
12. Mauldin, M.R.; McCollum, A.M.; Nakazawa, Y.J.; Mandra, A.; Whitehouse, E.R.; Davidson, W.; Zhao, H.; Gao, J.; Li, Y.; Doty, J.; et al. Exportation of Monkeypox Virus from the African Continent. *J. Infect. Dis.* **2022**, *225*, 1367–1376, doi:10.1093/infdis/jiaa559.
13. WHO Calls Emergency Meeting as Monkeypox Cases Top 100 in Europe Available online: https://www.reuters.com/world/europe/monkeypox-outbreak-europe-largest-ever-region-cases-cross-100-2022-05-20/ (accessed on 12 June 2022).
14. Unlikely Monkeypox Outbreak Will Lead to Pandemic, WHO Says Available online: https://www.reuters.com/business/healthcare-pharmaceuticals/unlikely-monkeypox-outbreak-will-lead-pandemic-says-who-2022-05-30/ (accessed on 12 June 2022).
15. Kelleher, S.R. CDC Raises Monkeypox Travel Alert to Level 2 Available online: https://www.forbes.com/sites/suzannerowankelleher/2022/06/07/cdc-raises-monkeypox-travel-alert-to-level-2/?sh=269eee1e3f93 (accessed on 12 June 2022).
16. Treatment Available online: https://www.cdc.gov/poxvirus/monkeypox/clinicians/treatment.html (accessed on 12 June 2022).
17. Available online: https://assets.publishing.service.gov.uk/government/ (accessed on 12 June 2022).
18. SIGA Receives Approval from the FDA for Intravenous (IV) Formulation of TPOXX® (Tecovirimat) Available online: https://investor.siga.com/news-releases/news-release-details/siga-receives-approval-fda-intravenous-iv-formulation-tpoxxr (accessed on 12 June 2022).
19. Gilchrist, K. Belgium Becomes First Country to Introduce Mandatory Monkeypox Quarantine as Global Cases Rise Available online: https://www.cnbc.com/2022/05/23/belgium-introduces-mandatory-monkeypox-quarantine-as-global-cases-rise.html (accessed on 12 June 2022).
20. Bahl, R. Monkeypox Vaccine: U.s. Orders 500,000 Jynneos Doses as Cases Rise Available online: https://www.healthline.com/health-news/monkeypox-vaccine-existing-vaccines-provide-strong-protection-one-fda-approved (accessed on 12 June 2022).
21. Toronto to Offer Monkeypox Vaccine Clinics Targeting High-Risk Communities. *CBC News*.
22. With, A.M. Push for Targeted Monkeypox Vaccine Rollout in France, Denmark Available online: https://www.rfi.fr/en/europe/20220525-push-for-targeted-monkeypox-vaccine-rollout-in-france-denmark (accessed on 12 June 2022).
23. Monkeypox: German Panel Recommends Vaccine for Risk Groups Available online: https://www.dw.com/en/monkeypox-german-panel-recommends-vaccine-for-risk-groups/a-62084728 (accessed on 12 June 2022).
24. UK. Health Authority Advises Self-Isolation for Monkeypox Infections Available online: https://www.thesundaily.my/world/uk-health-authority-advises-self-isolation-for-monkeypox-infections-NE9314681 (accessed on 12 June 2022).
25. Farias da Costa, V.C.; Oliveira, L.; de Souza, J. Internet of Everything (IoE) Taxonomies: A Survey and a Novel Knowledge-Based Taxonomy. *Sensors (Basel)* **2021**, *21*, 568, doi:10.3390/s21020568.
26. Auxier, B.; Anderson, M. Social Media Use in 2021 Available online: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/04/PI_2021.04.07_Social-Media-Use_FINAL.pdf (accessed on 12 June 2022).
27. Kavada, A. Social Media as Conversation: A Manifesto. *Soc. Media Soc.* **2015**, *1*, 205630511558079, doi:10.1177/2056305115580793.
28. Smith, A.; Brenner, J. Twitter Use 2012 Available online: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_Twitter_Use_2012.pdf (accessed on 12 June 2022).
29. Morgan-Lopez, A.A.; Kim, A.E.; Chew, RF; Ruddle, P. Predicting Age Groups of Twitter Users Based on Language and Metadata Features. *PLoS One* **2017**, *12*, e0183537, doi:10.1371/journal.pone.0183537.
30. Iqbal, M. Twitter Revenue and Usage Statistics (2022) Available online: https://www.businessofapps.com/data/twitter-statistics/ (accessed on 12 June 2022).
31. Lopez, C.E.; Gallemore, C. An Augmented Multilingual Twitter Dataset for Studying the COVID-19 Infodemic. *Soc. Netw. Anal. Min.* **2021**, *11*, 102, doi:10.1007/s13278-021-00825-0.

32. Gupta, R.K.; Vishwanath, A.; Yang, Y. COVID-19 Twitter Dataset with Latent Topics, Sentiments and Emotions Attributes. *arXiv [cs.CL]* **2020**, doi:10.48550/ARXIV.2007.06954.
33. Haq, E.-U.; Tyson, G.; Lee, L.-H.; Braud, T.; Hui, P. Twitter Dataset for 2022 Russo-Ukrainian Crisis. *arXiv [cs.SI]* **2022**, doi:10.48550/ARXIV.2203.02955.
34. Shevtsov, A.; Tzagkarakis, C.; Antonakaki, D.; Pratikakis, P.; Ioannidis, S. Twitter Dataset on the Russo-Ukrainian War. *arXiv [cs.SI]* **2022**, doi:10.48550/ARXIV.2204.08530.
35. Urchs, S.; Wendlinger, L.; Mitrovic, J.; Granitzer, M. MMoveT15: A Twitter Dataset for Extracting and Analysing Migration-Movement Data of the European Migration Crisis 2015. In Proceedings of the 2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE); IEEE, 2019; pp. 146–149.
36. Stemmer, M.; Parmet, Y.; Ravid, G. What Are IBD Patients Talking about on Twitter? In *ICT for Health, Accessibility and Wellbeing*; Springer International Publishing: Cham, 2021; pp. 206–220 ISBN 9783030942083.
37. Wijesiriwardene, T.; Inan, H.; Kursuncu, U.; Gaur, M.; Shalin, VL; Thirunarayan, K.; Sheth, A.; Arpinar, I.B. ALONE: A Dataset for Toxic Behavior among Adolescents on Twitter. In *Lecture Notes in Computer Science*; Springer International Publishing: Cham, 2020; pp. 427–439 ISBN 9783030609740.
38. Zangerle, E.; Pichl, M.; Gassler, W.; Specht, G. #nowplaying Music Dataset: Extracting Listening Behavior from Twitter. In Proceedings of the Proceedings of the First International Workshop on Internet-Scale Multimedia Management - WISMM '14; ACM Press: New York, New York, U.S.A., 2014.
39. Sech, J.; DeLucia, A.; Buczak, AL; Dredze, M. Civil Unrest on Twitter (CUT): A Dataset of Tweets to Support Research on Civil Unrest. In Proceedings of the Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020); Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 215–221.
40. Tekumalla, R.; Banda, J.M. A Large-Scale Twitter Dataset for Drug Safety Applications Mined from Publicly Existing Resources. *arXiv [cs. IR]* **2020**, doi:10.48550/ARXIV.2003.13900.
41. Mahase, E. Monkeypox: What Do We Know about the Outbreaks in Europe and North America? *BMJ* **2022**, *377*, o1274, doi:10.1136/bmj.o1274.
42. Pichl, M.; Zangerle, E.; Specht, G. Towards a Context-Aware Music Recommendation Approach: What Is Hidden in the Playlist Name? In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW); IEEE, 2015; pp. 1360–1365.
43. Jannach, D.; Kamehkhosh, I.; Lerche, L. Leveraging Multi-Dimensional User Models for Personalized next-Track Music Recommendation. In Proceedings of the Proceedings of the Symposium on Applied Computing - SAC '17; ACM Press: New York, New York, USA, 2017.
44. Ludewig, M.; Jannach, D. Evaluation of Session-Based Recommendation Algorithms. *User Model. User-adapt Interact.* **2018**, *28*, 331–390, doi:10.1007/s11257-018-9209-6.
45. Zangerle, E.; Chen, C.-M.; Tsai, M.-F.; Yang, Y.-H. Leveraging Affective Hashtags for Ranking Music Recommendations. *IEEE Trans. Affect. Comput.* **2021**, *12*, 78–91, doi:10.1109/taffc.2018.2846596.
46. Zangerle, E.; Pichl, M.; Hupfauf, B.; Specht, G. Can Microblogs Predict Music Charts? An Analysis of the Relationship between #nowplaying Tweets and Music Charts Available online: http://m.mr-pc.org/ismir16/website/articles/039_Paper.pdf (accessed on 1 July 2022).
47. Pichl, M.; Zangerle, E.; Specht, G. Understanding user-curated playlists on Spotify: A machine learning approach. *Int. J. Multimed. Data Eng. Manag.* **2017**, *8*, 44–59, doi:10.4018/ijmdem.2017100103.
48. Abboud, R.; Tekli, J. Integration of Nonparametric Fuzzy Classification with an Evolutionary-Developmental Framework to Perform Music Sentiment-Based Analysis and Composition. *Soft Comput.* **2020**, *24*, 9875–9925, doi:10.1007/s00500-019-04503-4.
49. Kaneshiro, B.; Ruan, F.; Baker, CW; Berger, J. Characterizing Listener Engagement with Popular Songs Using Large-Scale Music Discovery Data. *Front. Psychol.* **2017**, *8*, 416, doi:10.3389/fpsyg.2017.00416.
50. Zangerle, E.; Pichl, M.; Schedl, M. User Models for Culture-Aware Music Recommendation: Fusing Acoustic and Cultural Cues. *Trans. Int. Soc. Music Inf. Retr.* **2020**, *3*, 1–16, doi:10.5334/tismir.37.
51. Hridi, A.P. Mining User Personality from Music Listening Behavior in Online Platforms Using Audio Attributes, Clemson University, 2021.
52. Garcia, K.; Berton, L. Topic Detection and Sentiment Analysis in Twitter Content Related to COVID-19 from Brazil and the USA. *Appl. Soft Comput.* **2021**, *101*, 107057, doi:10.1016/j.asoc.2020.107057.
53. Gencoglu, O. Large-Scale, Language-Agnostic Discourse Classification of Tweets during COVID-19. *Mach. Learn. Knowl. Extr.* **2020**, *2*, 603–616, doi:10.3390/make2040032.
54. Al-Laith, A.; Alenezi, M. Monitoring People's Emotions and Symptoms from Arabic Tweets during the COVID-19 Pandemic. *Information (Basel)* **2021**, *12*, 86, doi:10.3390/info12020086.
55. Ng, HXL; Lee, R.K.-W.; Awal, MR I Miss You Babe: Analyzing Emotion Dynamics During COVID-19 Pandemic. In Proceedings of the Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 41–49.
56. Santarossa, S.; Rapp, A.; Sardinas, S.; Hussein, J.; Ramirez, A.; Cassidy-Bushrow, A.E.; Cheng, P.; Yu, E. Understanding the #longCOVID and #longhaulers Conversation on Twitter: Multimethod Study. *JMIR Infodemiology* **2022**, *2*, e31259, doi:10.2196/31259.

57. Lwin, M.O.; Sheldenkar, A.; Lu, J.; Schulz, P.J.; Shin, W.; Panchapakesan, C.; Gupta, R.K.; Yang, Y. The Evolution of Public Sentiments during the COVID-19 Pandemic: Case Comparisons of India, Singapore, South Korea, the United Kingdom, and the United States. *JMIR Infodemiology* **2022**, *2*, e31473, doi:10.2196/31473.
58. Mahase, E. Monkeypox: What Do We Know about the Outbreaks in Europe and North America? *BMJ* **2022**, *377*, o1274, doi:10.1136/bmj.o1274.
59. Bragazzi, N.L.; Khamisy-Farah, R.; Tsigalou, C.; Mahroum, N.; Converti, M. Attaching a Stigma to the LGBTQI+ Community Should Be Avoided during the Monkeypox Epidemic. *J. Med. Virol.* **2022**, doi:10.1002/jmv.27913.
60. Mahase, E. Seven Monkeypox Cases Are Confirmed in England. *BMJ* **2022**, *377*, o1239, doi:10.1136/bmj.o1239.
61. Ahsan, M.M.; Uddin, M.R.; Luna, S.A. Monkeypox Image Data Collection. *arXiv [eess.IV]* **2022**, doi:10.48550/ARXIV.2206.01774.
62. Heskin, J.; Belfield, A.; Milne, C.; Brown, N.; Walters, Y.; Scott, C.; Bracchi, M.; Moore, L.S.; Mughal, N.; Rampling, T.; et al. Transmission of Monkeypox Virus through Sexual Contact - A Novel Route of Infection. *J. Infect.* **2022**, doi:10.1016/j.jinf.2022.05.028.
63. Yoo, J.-H. Once Bitten, Twice Shy: Our Attitude towards Monkeypox. *J. Korean Med. Sci.* **2022**, *37*, e188, doi:10.3346/jkms.2022.37.e188.
64. Bisanzio, D.; Reithinger, R. Keep Calm and Carry on: Projected Case Burden and Duration of the 2022 Monkeypox Outbreak in Non-Endemic Countries. *bioRxiv* 2022.
65. Rao, A.K.; Petersen, B.W.; Whitehill, F.; Razeq, J.H.; Isaacs, S.N.; Merchlinsky, M.J.; Campos-Outcalt, D.; Morgan, R.L.; Damon, I.; Sánchez, P.J.; et al. Use of JYNNEOS (Smallpox and Monkeypox Vaccine, Live, Nonreplicating) for Preexposure Vaccination of Persons at Risk for Occupational Exposure to Orthopoxviruses: Recommendations of the Advisory Committee on Immunization Practices - United States, 2022. *MMWR Morb. Mortal. Wkly. Rep.* **2022**, *71*, 734–742, doi:10.15585/mmwr.mm7122e1.
66. Vandenbogaert, M.; Kwasiborski, A.; Gonofio, E.; Descorps-Declère, S.; Selekon, B.; Nkili Meyong, AA; Ouilibona, R.S.; Gessain, A.; Manuguerra, J.-C.; Caro, V.; et al. Nanopore Sequencing of a Monkeypox Virus Strain Isolated from a Pustular Lesion in the Central African Republic. *Sci. Rep.* **2022**, *12*, 10768, doi:10.1038/s41598-022-15073-1.
67. Miura, F.; van Ewijk, C.E.; Backer, J.A.; Xiridou, M.; Franz, E.; de Coul, E.O.; Brandwagt, D.; van Cleef, B.; van Rijckevorsel, G.; Swaan, C.; et al. The Incubation Period for Monkeypox Cases Confirmed in the Netherlands, May 2022. *bioRxiv* 2022.
68. Dashraath, P.; Nielsen-Saines, K.; Mattar, C.; Musso, D.; Tambyah, P.; Baud, D. Guidelines for Pregnant Individuals with Monkeypox Virus Exposure. *Lancet* **2022**, doi:10.1016/S0140-6736(22)01063-7.
69. Malik, A.A.; Winters, M.S.; Omer, S.B. Attitudes of the US General Public towards Monkeypox. *bioRxiv* 2022.
70. Alcolea-Medina, A.; Charalampous, T.; Snell, L.B.; Aydin, A.; Alder, C.; Maloney, G.; Bryan, L.; Nebbia, G.; Douthwaite, S.; Neil, S.; et al. Novel, Rapid Metagenomic Method to Detect Emerging Viral Pathogens Applied to Human Monkeypox Infections. *SSRN Electron. J.* **2022**, doi:10.2139/ssrn.4132526.
71. Bellinato, F.; Gisondi, P.; Girolomoni, G. Monkeypox Virus Infection: What Dermatologist Needs to Know? *J. Eur. Acad. Dermatol. Venereol.* **2022**, doi:10.1111/jdv.18299.
72. Thakur, N. MonkeyPox2022Tweets: The First Public Twitter Dataset on the 2022 MonkeyPox Outbreak. *Preprints* 2022.
73. Thakur, N. MonkeyPox2022Tweets: The First Public Twitter Dataset on the 2022 MonkeyPox Outbreak Available online: http://dx.doi.org/10.21227/16CA-C879.
74. Thakur, N. MonkeyPox2022Tweets: The First Public Twitter Dataset on the 2022 MonkeyPox Outbreak Available online: https://zenodo.org/record/6760926#.YrobZXbMLIU.
75. Thakur, N. Twitter Dataset on the 2022 MonkeyPox Outbreak Available online: https://www.kaggle.com/datasets/thakurnirmalya/monkeypox2022tweets (accessed on 1 July 2022).
76. Privacy Policy Available online: https://twitter.com/en/privacy/previous/version_15 (accessed on 12 June 2022).
77. Developer Agreement and Policy Available online: https://developer.twitter.com/en/developer-terms/agreement-and-policy (accessed on 12 June 2022).
78. Hydrator: Turn Tweet IDs into Twitter JSON & CSV from Your Desktop! Available online: https://github.com/DocNow/hydrator (accessed on 12 June 2022).
79. Tekumalla, R.; Banda, J.M. Social Media Mining Toolkit (SMMT). *Genomics Inform.* **2020**, *18*, e16, doi:10.5808/GI.2020.18.2.e16.
80. Twarc: A Command Line Tool (and Python Library) for Archiving Twitter JSON Available online: https://github.com/DocNow/twarc (accessed on 12 June 2022).
81. Standard Search API Available online: https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets (accessed on 12 June 2022).
82. Mierswa, I.; Wurst, M.; Klinkenberg, R.; Scholz, M.; Euler, T. YALE: Rapid Prototyping for Complex Data Mining Tasks. In Proceedings of the Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06; ACM Press: New York, New York, U.S.A., 2006.
83. RapidMiner GmbH Search Twitter - RapidMiner Documentation Available online: https://docs.rapidminer.com/latest/studio/operators/data_access/applications/twitter/search_twitter.html (accessed on 12 June 2022).
84. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018, doi:10.1038/sdata.2016.18.

doi:10.20944/preprints202206.0172.v3