doi:10.20944/preprints202205.0238.v1

Article

An Exploratory Study of Tweets about the SARS-CoV-2 Omicron Variant: Insights from Sentiment Analysis, Language Interpretation, Source Tracking, Type Classification, and Embedded URL Detection

Nirmalya Thakur [1]* and Chia Y. Han [2]

[1] Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221-0030, U.S.A.; thakurna@mail.uc.edu
[2] Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221-0030, U.S.A.; han@ucmail.uc.edu
* Correspondence: thakurna@mail.uc.edu

Abstract: This paper presents the findings of an exploratory study on the continuously generating Big Data on Twitter related to the sharing of information, news, views, opinions, ideas, knowledge, feedback, and experiences about the COVID-19 pandemic, with a specific focus on the Omicron variant, which is the globally dominant variant of SARS-CoV-2 at this time. A total of 12028 tweets about the Omicron variant were studied, and the specific characteristics of tweets that were analyzed include - sentiment, language, source, type, and embedded URLs. The findings of this study are manifold. First, from sentiment analysis, it was observed that 50.5% of tweets had the 'neutral' emotion. The other emotions - 'bad', 'good', 'terrible', and 'great' were found in 15.6%, 14.0%, 12.5%, and 7.5% of the tweets, respectively. Second, the findings of language interpretation showed that 65.9% of the tweets were posted in English. It was followed by Spanish or Castillian, French, Italian, Japanese, and other languages, which were found in 10.5%, 5.1%, 3.3%, 2.5%, and <2% of the tweets, respectively. Third, the findings from source tracking showed that "Twitter for Android" was associated with 35.2% of tweets. It was followed by "Twitter Web App", "Twitter for iPhone", "Twitter for iPad", "TweetDeck", and all other sources that accounted for 29.2%, 25.8%, 3.8%, 1.6%, and <1% of the tweets, respectively. Fourth, studying the type of tweets revealed that retweets accounted for 60.8% of the tweets, it was followed by original tweets and replies that accounted for 19.8% and 19.4% of the tweets, respectively. Finally, in terms of embedded URL analysis, the most common URLs embedded in the tweets were found to be twitter.com, which was followed by biorxiv.org, nature.com, wapo.st, nzherald.co.nz, recvprofits.com, science.org, and other URLs.

Keywords: COVID-19; SARS-CoV-2; Omicron; Twitter; tweets; sentiment analysis; big data; Natural Language Processing; Data Science; Data Analysis

1. Introduction

An outbreak of an unknown respiratory disease started in December 2019 in the seafood market in Wuhan, China, infecting about 66% of the people at the market. Very soon, more people in different parts of China got infected by the same disease. This prompted an investigation from the healthcare and medical sectors, and very soon, it was concluded that a novel coronavirus was responsible for this disease. This novel coronavirus was named severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2, 2019-nCoV) as it was observed to have a high homology of about 80% with SARS-CoV [1]. The disease that humans suffer from after getting infected by this virus is known as COVID-19 [2]. The Chinese government implemented multiple measures to reduce the spread of the virus. However, the virus rapidly spread across China and soon started spreading in different

 doi:10.20944/preprints202205.0238.v1

countries of the world. COVID-19 was declared a pandemic by the World Health Organization (WHO) on March 11, 2021 [4].

At the time of writing this paper, globally, there have been 520,628,068 cases of COVID-19 with 6,287,207 deaths [5]. The SARS-CoV-2 virus mainly attacks the respiratory system of humans, although infections in other organs of the body have also been reported in some cases. The symptoms, as reported from the initial studies of the cases from Wuhan, China, include fever, dry cough, dyspnea, headache, dizziness, exhaustion, vomiting, and diarrhea. The report also mentions that not everyone has the same symptoms, and the nature and intensity of the symptoms vary from person to person [6]. When the genetic sequence of a virus changes, it is said to have mutated. Genomes of a virus that are different from each other in terms of their genetic sequence are called variants. Variants that differ in terms of the phenotype are known as strains [7]. On January 10, 2020, the genome sequences of SARS-CoV-2 were made publicly available by the Global Influenza Surveillance and Response System (GISAID), which is a primary source on a global scale for open access to the genomic data of influenza viruses [8]. Since then, the database of GISAD has made more than 5 million genetic sequences of SARS-CoV-2 from 194 countries and territories publicly available for research [9]. In an attempt to prioritize research related to COVID-19, the WHO has made a conscious effort to classify the variants of SARS-CoV-2 into three categories. These include variants of concern (VOCs), variants of interest (VOIs), and variants under monitoring (VUMs).

Since the initial cases in December 2019, the SARS-CoV-2 virus has undergone multiple mutations, and as a result, several variants have been detected in different parts of the world. Some of these include – Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Epsilon (B.1.427 B.1.429), Eta (B.1.525), Iota (B.1.526), Kappa (B.1.617.1), Zeta (P.2), Mu (B.1.621, B.1.621.1), and Omicron (B.1.1.529, BA.1, BA.1.1, BA.2, BA.3, BA.4 and BA.5) [10]. Out of all these variants, the Omicron variant, first detected on November 24, 2021, was classified as a VOC. by WHO on November 26, 2021 [11]. Soon thereafter, the Omicron variant became the globally dominant form of SARS-CoV-2 [12]. The Omicron spike protein contains 30 mutations and has been reported to be the most immune evasive VOC. of SARS-CoV-2 and is highly resistant against antibody-mediated neutralization [13,14]. Despite the development of vaccines [15] and various forms of treatments [16], a recent report from WHO states that the Omicron variant accounts for 86% of the infections caused by SARS-CoV-2 on a global scale [17]. In another recent report, WHO mentioned that the cases due to the Omicron variant were "off the charts" as global infections due to the SARS-CoV-2 variant set new records [18]. Currently, some of the countries that have recorded the most cases due to the SARS-CoV-2 Omicron variant include – United Kingdom (1,138,814 cases), USA (945,470 cases), Germany (245,120 cases), Denmark (218,106 cases), France (110,959 cases), Canada (92,341 cases), Japan (71,056 cases), India (56,125 cases), Australia (46,576 cases), Sweden (43,400 cases), Israel (39,908 cases), Poland (33,436 cases), and Brazil (32,880 cases) [19].

In today's world, where the internet virtually connects people in different geographic regions, social media has been considered an "integral vehicle" of people's lives and as an "online community" for communication, information, news, views, opinions, perspectives, ideas, knowledge, feedback, and experiences related to pandemics, epidemics, viruses, and diseases [20-24]. Out of all the social media platforms, Twitter is popular amongst all age groups [25], and there are about 192 million daily active users on Twitter [26]. Twitter has been highly popular amongst healthcare researchers, epidemiologists, medical practitioners, data scientists, and computer science researchers for studying, analyzing, modeling, and interpreting social media communications related to pandemics, epidemics, viruses, and diseases such as Ebola [27], E-Coli [28], Dengue [29], Human papillomavirus (HPV) [30], Middle East Respiratory Syndrome (MERS) [31], Measles [32], Zika virus [33], H1N1 [34], influenza-like illness [35], swine flu [36], flu [37], Cholera [38], Listeriosis [39], cancer [40], Liver Disease [41], Inflammatory Bowel Disease [42], kidney disease [43], lupus [44], Parkinson's [45], Diphtheria [46], and West Nile virus [47].

The outbreak of the SARS-CoV-2 served as a "catalyst" towards increasing the usage of Twitter [48-51] for conversations on a wide range of topics in this regard resulting in tremendous amounts of Big Data. Some of the most popular use cases of Twitter during this pandemic, as reported in recent works in this field, included - (1) sharing of symptoms, information, and experiences as reported by frontline works and people who were infected with the virus [52]; (2) providing suggestions, opinions, and recommendations to reduce the spread of the virus [53]; (3) updates on vaccine development, clinical trials, and other forms of treatment [54]; guidelines mandated by various policy-making bodies such as mask mandate, social distancing, etc. that were required to be followed by members in specific geographic regions of the world under the authority of the associated policy-making bodies [55]; (4) dissemination of misinformation such as use of certain drugs or forms of treatment that have not been tested or have not undergone clinical trials [56]; (5) conspiracy theories such as considering 5G technologies responsible for the spread of COVID-19 which eventually led to multiple 5G towers being burnt down in the United Kingdom [57]; and (6) studying public oppositions to available vaccines in different parts of the world [58]. In addition to these works related to the use of Twitter during COVID-19, there have been some other works in this field (as discussed in detail in Section 2). However, none of these works took a comprehensive approach towards drawing insights related to Twitter usage in the context of tweets related to SARS-CoV-2 with a specific focus on the SARS-CoV-2 Omicron variant. To address this research challenge, in this work, a total of 12,028 recent tweets about the SARS-CoV-2 Omicron variant posted publicly on Twitter from May 5, 2022, to May 12, 2022, were studied, analyzed, and interpreted to perform the following. These tweets had a combined reach, impressions, retweets, and favorites of 149,500,959, 226,603,833, 1,053,869, and 3,427,976, respectively.

(1) To detect the sentiment of the associated tweets in terms of 'great', 'good', 'neutral', 'bad', and 'terrible'. The findings show that a majority of the tweets (50.5%) had a 'neutral' emotion, which was followed by the emotional states of 'bad', 'good', 'terrible', and 'great' found in 15.6%, 14.0%, 12.5%, and 7.5% of the tweets, respectively.

(2) To track the source, such as the android phone or tablet, iPhone, Web App, etc., that was used to post the tweets. The findings show that 35.2% of the Tweets were posted from an android source, which was followed by the Twitter Web App, iPhone, iPad, Tweetdeck, and other sources, which accounted for 29.2%, 25.8%, 3.8%, 1.6%, and <1% of the tweets, respectively.

(3) To interpret the language of the associated tweets by using language recognition. The findings show that 65.9% of the tweets were posted in English, which was followed by Spanish or Castillian (10.5%), French (5.1%), Italian (3.3%), Japanese (2.5%), and other languages that accounted for <2% of the tweets.

(4) To classify the tweets by Tweet Type in terms of original tweets, retweets, and replies. The findings show that the majority of the tweets (60.8%) were retweets which was followed by original tweets (19.8%) and replies (19.4%).

(5) To identify the most common URLs embedded in the associated tweets. The most common URL embedded in the tweets was found to be twitter.com, which was followed by biorxiv.org, nature.com, wapo.st, nzherald.co.nz, recvprofits.com, science.org, and a few other sources.

This paper is presented as follows. In Section 2, a review of recent works in this field is presented. The methodology is discussed in Section 3. Section 4 discusses the results. The conclusion is presented in Section 5, which summarizes the scientific contributions of this study and outlines the scope for future work in this field. It is followed by the references section.

2. Literature Review

In this section, a review is presented of the recent works in this field that focused on studying social media behavior with a specific focus on Twitter since the outbreak of SARS-CoV-2. Shamrat et al. [59] developed a kNN-based machine learning classifier to classify tweets related to COVID-19 into three classes – 'positive', 'negative', and 'neutral'. The study specifically focused on filtering tweets related to COVID-19 vaccines, and thereafter this algorithm was applied to the filtered tweets for analysis. Sontayasara et al. [60] used the support vector machine (SVM) classifier to develop an algorithm for sentiment analysis. The algorithm was tested on tweets where people communicated their plans to visit Bangkok during the pandemic and how those plans were affected. This classifier was able to classify the tweets into three classes of sentiments - 'positive', 'negative', and 'neutral'. Asgari-Chenaghlu et al. [61] developed an approach to detect the trending topics and major concerns related to the COVID-19 pandemic as expressed by people on Twitter. Amen et al. [62] proposed a framework that applied a directed acyclic graph model on Tweets related to COVID-19 to detect any anomaly events. The work of Liu et al. [63] involved developing an approach to study tweets about COVID-19 that involved the Centers for Disease Control and Prevention (C.D.C.). The objective of this study was to detect public perceptions such as concerns, attention, expectations, etc., related to the guidelines of the C.D.C. regarding COVID-19. Al-Ramahi et al. [64] developed a methodology to filter and study tweets posted between January 1, 2020, and October 27, 2020, where people expressed their opposing views towards wearing masks to reduce the spread of COVID-19. Jain et al. [65] proposed a methodology to analyze tweets related to COVID-19 that could assign an influence score to the associated users who posted these tweets. The objective of this study was to identify influential users on Twitter who posted about COVID-19. Madani et al. [66] developed a random forest-based classifier to detect tweets about COVID-19 that contained fake news. The classifier achieved a performance accuracy of 79%.

Shokoohyar et al. [67] proposed a system to study tweets where people expressed their opinions regarding the lockdown in the United States on account of COVID-19. Chehal et al. [68] developed a software using Python and R to analyze the mindset of Indians as expressed in their tweets during the two nationwide lockdowns that were implemented by the Indian government on account of COVID-19. Glowacki et al. [69] developed a systemic approach to identify and study tweets related to COVID-19 where Twitter users discussed addiction issues. Selman et al. [70]'s study focused on studying tweets where Twitter users reported their relative, friend, or acquaintance passing away from COVID-19. The study specifically focused on patients who were reported to have been alone at the time of their death. The work of Koh et al. [71] aimed to identify tweets using specific keywords where Twitter users communicated about feelings of loneliness during COVID-19. The authors tested their approach on a total of 4492 tweets. Mackey et al. [72]'s work focused on filtering and investigating tweets related to COVID-19 where people self-reported their symptoms, access to testing sites, and recovery status. In [73], the authors focused on studying tweets related to COVID-19 to understand the anxiety and panic buying behavior with a specific focus on buying toilet paper during this pandemic. The work involved specific inclusion criteria for the tweets, and a total of 4081 tweets were studied. In addition to the above, there have been multiple datasets [73-76] that have been developed by researchers by filtering tweets related to COVID-19 since the beginning of the pandemic.

As can be seen from these works involving studying social media behavior and user characteristics on Twitter during COVID-19, while there have been several innovations and advancements made in this field, the following limitations exist in these works:

1. Most of these works used approaches to detect tweets that contained one or more keywords, hashtags, or phrases such as "COVID-19", "coronavirus", "SARS-CoV-2", "covid", "corona," etc. but none of these works focused on including one or more keywords directly related to the SARS-CoV-2 omicron variant to include the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 May 2022 doi:10.20944/preprints202205.0238.v1

associated tweets. As the SARS-CoV-2 Omicron variant is now responsible for most of the COVID-19 cases globally, the need in this context is to filter tweets that contain one or more keywords, hashtags, or phrases related to this variant.

2. The works on sentiment analysis [59, 60] focused on the proposal of new approaches to detect the sentiment associated with tweets; however, the categories for classification of the associated sentiment were only 'positive', 'negative', and 'neutral'. In a realistic scenario, there can be different kinds of 'positive' emotions, such as 'good' and 'great'. Similarly, there can be different kinds of 'negative' emotions, such as 'bad' and 'terrible'. The existing works cannot differentiate between these kinds of positive (or negative) emotions. Therefore, the need in this context is to expand the levels of sentiment classification to include the different kinds of positive and negative emotions.
3. While there have been multiple innovations in this field of Twitter data analysis - such as detecting trending topics [61], anomaly events [62], public perceptions towards C.D.C. [63], and views towards not wearing masks [64], just to name a few, there has been minimal work related to quantifying and ranking the associated insights.
4. The number of tweets that were included in previous studies (such as 4081 tweets in [73] and 4492 tweets in [71]) comprises a very small percentage of the total number of tweets that have been posted related to COVID-19 since the beginning of the outbreak. Therefore, the need in this context is to include more tweets in the studies.

The work proposed in this paper aims to explore the intersections of Big Data mining, Natural Language Processing, Data Science, Information Retrieval, Machine Learning, and their related areas to address the above-mentioned needs. The methodology is outlined in the next section.

3. Methodology

This section presents the methodology of the work proposed in this paper. The work primarily involves drawing insights by mining tweets. Therefore the privacy policy [78] and developer agreement and policy [79] of Twitter were studied at first. The privacy policy of Twitter [78] states – *"Twitter is public and Tweets are immediately viewable and searchable by anyone around the world"*. To add, the Twitter developer agreement and policy [79] defines tweets as *"public data"*. Therefore, based on the terms and conditions mentioned in these two policies at the time of writing of this paper, it can be concluded that performing this data analysis and drawing insights from tweets is compliant and adheres to both these policies.

The step-by-step process that was followed is represented in the flowchart shown in Figure 1. As can be seen from Figure 1, the first step was to perform the Big Data mining to collect these tweets from Twitter. The Twitter search API has a 7-day limit on the tweets that can be searched [80]; in other words, tweets posted more than 7-days ago cannot be searched and studied. So, a total of 12028 recent tweets about the SARS-CoV-2 Omicron variant were studied that were posted on Twitter between May 12, 2022 (the most recent date at the time of data collection) and May 5, 2022 (the date up to which tweets could be searched in compliance with the Twitter search API guidelines). This data mining process was performed by filtering tweets from Twitter in this date range that contained the "omicron" keyword or hashtag. A list of 7 tweets (selected at random) out of this corpus of 12028 tweets is shown in Table 1. As can be seen from this random collection of tweets, there were different phrases that constituted referring to the SARS-CoV-2 omicron variant. These phrases included – "first case of omicron variant", "approaching its Omicron peak", "Omicron is a stone cold killer", "case of Omicron sub-variant", "new omicron subvariants", "the highly transmissible omicron variant", "sinus pain of omicron". Even though these phrases are different, the keyword "omicron" is present in all these phrases. This helps to uphold the relevance of performing the data mining and data collection of relevant tweets by using this keyword. The process of data mining that was followed was

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 May 2022 doi:10.20944/preprints202205.0238.v1

not case sensitive, so the tweets containing the "omicron" keyword (or hashtag) as "omicron" or "Omicron" or "OMICRON" or any other order of capitalization of these alphabets were also filtered, and these specific cases did not need to be specifically mentioned as separate or different keywords (or hashtags).

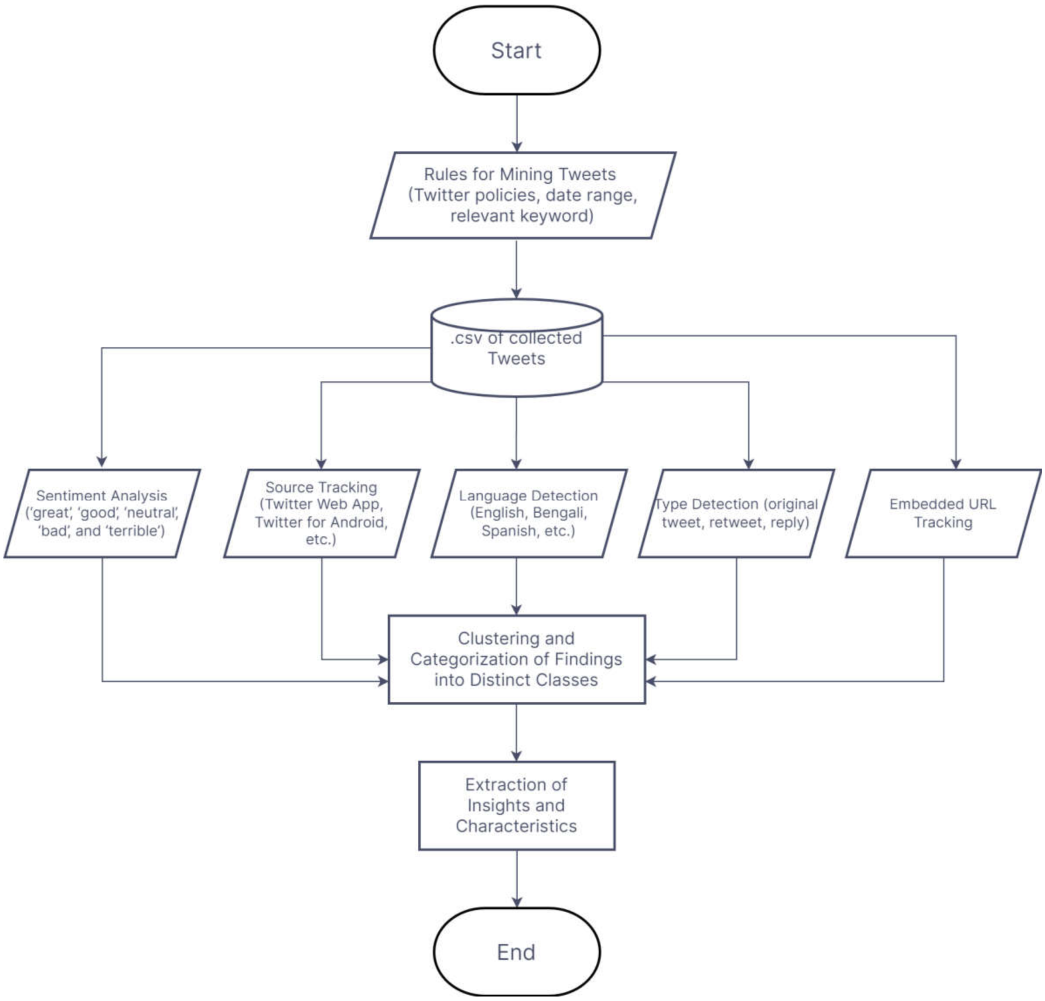

Figure 1. Flowchart-based representation of the methodology that was followed.

Table 1. A random selection of 7 tweets from the corpus of 12028 tweets that were studied in this work.

Tweet No.	Tweet Text
Tweet #1	N. Korea reports first case of omicron variant of COVID-19
Tweet #2	Vermont—the most vaccinated state in the U.S.—now has 12/14 counties in the "high community level" and its hospital admissions are approaching its Omicron peak. Its experience reaching this level highlights some of the limitations of the C.D.C. guidance.
Tweet #3	Quebec's 2022 covid death toll has officially surpassed the 2021 death toll. We matched the death toll for the entire year of 2021 in just the first 4 months of 2022. Omicron has been one of the deadliest phases of this pandemic. Mild? No. Omicron is a stone cold killer.
Tweet #4	N.I.H. has detected the first case of Omicron sub-variant BA.2.12.1. This new sub-variant is causing increasing number of cases in different countries.
Tweet #5	Three new omicron sub-variants are detected in Australia! Vaccine is virus, there are hundreds of various now and more will discover!! COVID Virus is C.C.P. #bioweapon!!
Tweet #6	Scientists around the world are investigating how a dwindling number of people have managed to dodge the coronavirus for more than two years, even after the highly transmissible omicron variant drove a record-shattering surge in cases this winter.
Tweet #7	The burning sinus pain of omicron is awful. It's like the feeling of a very sore throat but up high in my sinuses, right at the base of my nasal bone.

For all these tweets (available as a .csv file after the data mining process was completed), the specific insights that were studied, computed, interpreted, analyzed, quantified, and ranked in the next steps included the sentiment, source tracking, detecting the language of the tweet, inferring the type of tweet, and detecting the URLs embedded in the tweets. The process of sentiment analysis involved tokenization and lemmatization of the text of the tweet and classifying it into various sentiments such as 'great', 'good', 'neutral', 'bad', and 'terrible'. While there can be multiple ways of performing sentiment analysis, this process of sentiment analysis was selected as multiple sentiment classes can be created, and the resultant classification is neither a binary classification between 'positive' and 'negative' sentiments nor is it a 3 level classification between 'positive', 'negative' and 'neutral' sentiment classes, which have been the limitations in prior works in this field (discussed in Section 2). The source tracking aspect of the study involved tracking the publicly available "source label" [81] that Twitter associates with each tweet. This source level has different categories, which include - Twitter Web App, Twitter for Android, Twitter for iPhone, and several more. Table 2 explains the condition for the assignment of four such "source labels" by Twitter.

Table 2. Four Tweet Source Labels and the conditions for assignment of these labels by Twitter.

Tweet Source Label	Condition for Assignment
Twitter Web App	The tweet was posted by visiting the official website of Twitter [82]
Twitter for Android	The tweet was posted using the Twitter app for Android operating systems which is available for free download on the Google Playstore [83]
Twitter for iPhone	The tweet was posted using the Twitter app for iPhone, which is available for free download on the App Store [84]
TweetDeck	The tweet was posted by using TweetDeck, a social media dashboard application for the management of Twitter accounts [85]

The Twitter platform allows its users to tweet in any of the 34 languages supported by Twitter [86]. Each of these languages is assigned a unique 2-letter code in the Twitter developer API, which helps to uniquely identify the associated language. Some examples include "en" for English, "ar" for Arabic, "bn" for Bengali, "cz" for Czech, "da" for Danish, "de" for German, "el" for Greek, "es" for Spanish, "fa" for Persian, and so on.

 doi:10.20944/preprints202205.0238.v1

Inferring the type of tweet involved tracking the publicly available information about a tweet on Twitter that mentions whether it is an original tweet or a retweet (an original tweet that has been re-posted) or a reply (response to an original tweet or a retweet). Finally, detecting the URL embedded in a tweet involved processing the text associated with a tweet to detect any URLs that may have been included in the tweet.

After detecting these characteristics and features from the set of 12028 tweets, in the next step, they were grouped together into distinct classes for quantification, categorization, and ranking. This helped to deduce multiple insights from each category of the results. For instance, in the tweet type category, this analysis helped to deduce the percentage of original tweets, retweets, and replies, which further helped in ranking these specific classes of results in the tweet type category. The ranking process helped to determine which of these respective classes constituted the maximum occurrence in each category. The above-mentioned functionalities for data mining, sentiment analysis, source tracking, language interpretation, type detection, and embedded URL analysis of Tweets were performed in a collective manner as per the flowchart by using Social Bearing, a research tool for performing Twitter research [87]. It was developed by Tom Elliott, and the tool was made available for the public starting on January 1, 2015. The tool uses multiple JavaScript, Jquery, JSON, text processing, and text analysis libraries and algorithms in combination as well as in standalone form for performing Big Data Mining, Data Analysis, Information Processing, and Information Retrieval on tweets (obtained based on keyword or hashtag search from Twitter) while adhering to the privacy policy, developer agreement, and developer policies of Twitter [78,79]. The results that were obtained are discussed in Section 4.

4. Results and Discussions

This section presents and discusses the results obtained upon the development and implementation of the proposed methodology on the set of 12028 tweets about the SARS-CoV-2 omicron variant. The results are shown in Figures 2-6. Figure 2 shows the results of sentiment analysis. The specific categories into which the sentiment of a tweet was classified comprised 'great', 'good', 'neutral', 'bad', and 'terrible'. As can be seen from Figure 2, the 'neutral' emotion was present in a majority of the tweets (50.5% of the total tweets). It was followed by tweets that had the 'bad' (15.5% of the total tweets) and 'good' (14.0% of the total tweets) emotions associated with them. These respective sentiment categories were followed by the sentiment categories of 'terrible' (12.5% of the total tweets) and 'great' (7.5% of the total tweets).

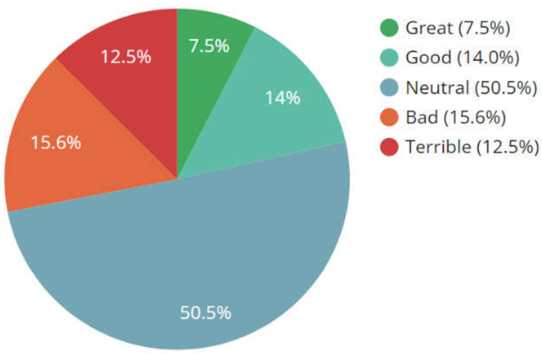

Figure 2. Results of sentiment analysis performed on tweets about the SARS-CoV-2 omicron variant.

The results of the Tweet Type detection are shown in Figure 3. This process involved detection, categorization, and ranking of the tweets into original tweets, retweets, and replies. In the index shown in this Figure, original tweets are referred to as "tweets". As can be seen from this Figure, retweets comprised a majority (60.8%) of all the tweets about the SARS-CoV-2 omicron variant. It was followed by original tweets (19.8% of the total tweets) and replies (19.4% of the total tweets).

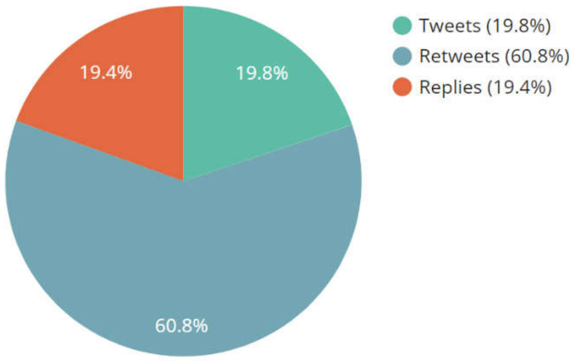

Figure 3. Results of Tweet type detection performed on tweets about the SARS-CoV-2 omicron variant.

Figure 4 shows the results of tweet source detection as publicly shown by the Twitter platform. Here, "Android" refers to "Twitter for Android", "iPhone" refers to "Twitter for iPhone", and "iPad" refers to "Twitter for iPad" (meanings of these categories are mentioned in Table 2). The total number of sources in this corpus of 12028 tweets was observed to be more than 10; the top 10 sources are listed in the index provided in this Figure for clarity and readability. As can be seen from this Figure, "Twitter for Android" accounted for the most number of tweets (35.2% of the total tweets). It was followed by "Twitter Web App" (29.2% of the total tweets), "Twitter for iPhone" (25.8% of the total tweets), "Twitter for iPad" (3.8% of the total tweets), and "TweetDeck" (1.6% of the total tweets). "TweetDeck" was followed by a few other sources, which accounted for less than 1% of the total tweets.

The results of the language detection and ranking of the used languages are shown in Figure 5. The total number of languages detected from all the 12028 tweets was observed to be more than 10; the top 10 languages are listed in the index provided in this Figure for clarity and readability. As can be seen from this Figure, more than a majority (65.9%) of the tweets were written in English. In the second place was the Spanish or Castilian (10.5% of the total tweets), which was followed by French (5.1% of the total tweets), Italian (3.3% of the total tweets), Japanese (2.5% of the total tweets), and a few other sources that accounted for less than 2% of the total tweets.

The results of the detection of the embedded URLs and the ranking of the associated clusters are shown in Figure 6. The total number of different URLs in this corpus of 12028 tweets was observed to be more than 10; the top 10 embedded URLs are listed in the index provided in this Figure for clarity and readability. As can be seen from Figure 6, the URL – twitter.com comprised the highest count. This can be attributed to the fact that retweets comprised a significant percentage of the tweets about the SARS-CoV-2 omicron variant. It was followed by the URLs - biorxiv.org, nature.com, wapo.st, nzherald.co.nz, recvprofits.com, science.org, bit.ly, YouTube (youtu.be is a shortened version of YouTube URLs [81]), and sciencedirect.com.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 May 2022 doi:10.20944/preprints202205.0238.v1

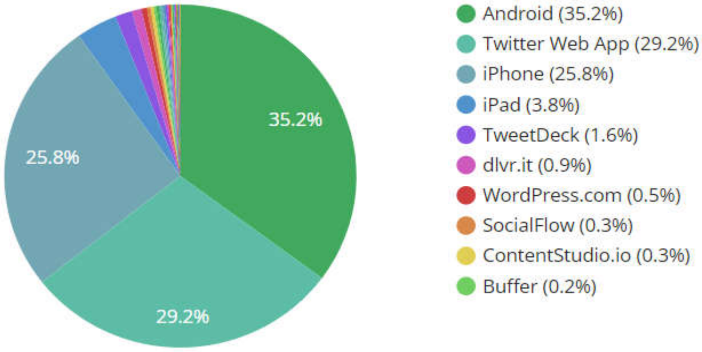

Figure 4. Results of Tweet source detection performed on tweets about the SARS-CoV-2 omicron variant.

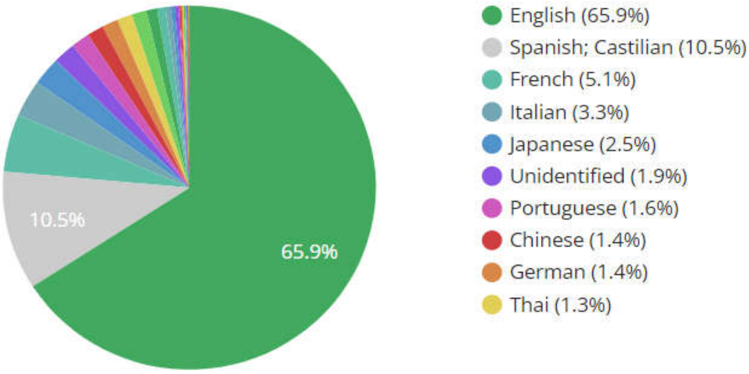

Figure 5. Results of language detection performed on tweets about the SARS-CoV-2 omicron variant.

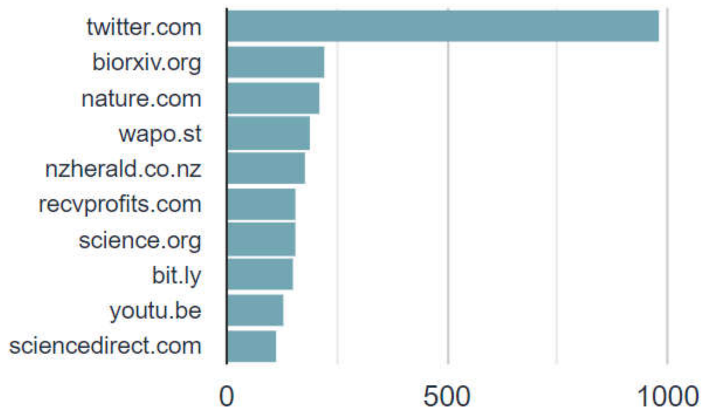

Figure 6. Results of embedded URL detection performed on tweets about the SARS-CoV-2 omicron variant.

In addition to the above, a keyword-based word cloud analysis and a hastag-based word cloud analysis were performed on this set of 12028 tweets about the SARS-CoV-2 omicron variant. The results are shown in Figures 7 and 8, respectively. In each of these word clouds, a different font color has been used to identify a keyword and hashtag, respectively. The font size of each of these words and hashtags is directly proportional to the frequency of the same. Or in other words, the keyword (or hashtag) that has the highest frequency has the largest font size, and the keyword (or hashtag) that has the lowest frequency has the smallest font size. Those keywords (or hashtags) which had a very low frequency and would subsequently have an extremely small font size were not included in the results shown to avoid cluttering and to enhance the readability of these Figures. As can be seen from both these figures, the "omicron" keyword was the most frequent

 doi:10.20944/preprints202205.0238.v1

keyword, and similarly, the "#omicron" hashtag was the most frequent hashtag across all these tweets. This observation further helps to support the proposed approach for searching tweets by using "omicron" as the keyword or hashtag to be searched, which was outlined at the beginning of this section.

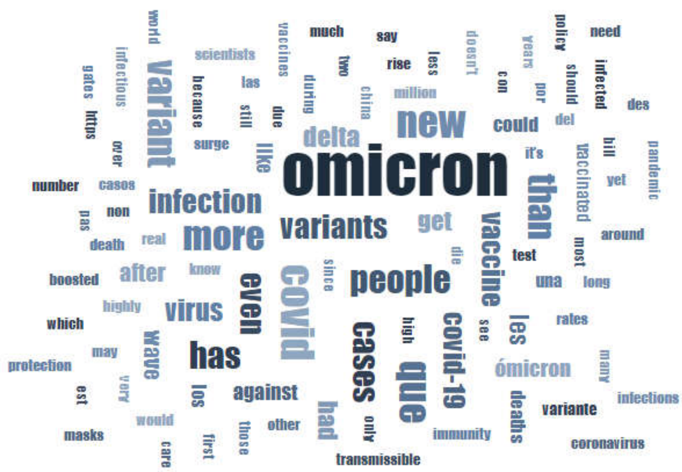

Figure 7. Results of word cloud analysis of the keywords present in the tweets about the SARS-CoV-2 omicron variant.

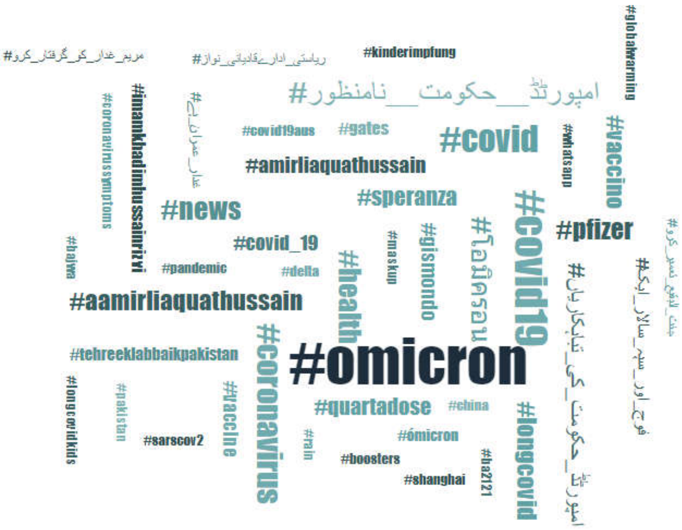

Figure 8. Results of word cloud analysis of the hashtags present in the tweets about the SARS-CoV-2 omicron variant.

In addition to these findings, a discussion is presented next that upholds how this work addresses the four distinct needs in this field of research that were identified after a comprehensive review of recent works (presented in Section 2).

1. The previous works in this field proposed approaches to filter tweets based on one or more keywords, hashtags, or phrases such as "COVID-19", "coronavirus", "SARS-CoV-2", "covid", "corona" but did not contain any keyword or phrase specifically related to the omicron variant. So, those approaches for tweet searching or tweet filtering might not be applicable to the tweets posted about the omicron variant unless the Twitter user specifically mentions something like "COVID-19 omicron variant" or "SARS-CoV-2 Omicron variant" in their tweets. As can be seen from the random collection of tweets presented in Table 1, there were multiple instances when the Twitter users did not use keywords, hashtags, or phrases such as "COVID-19", "coronavirus", "SARS-CoV-2", "covid", "corona" along with the keyword or hashtag – "omicron". Thus the need is to develop an approach to specifically mine tweets

posted about the omicron variant. This work addresses this need by proposing a methodology that searches tweets based on the presence of "omicron" either as a keyword or as a hashtag. The effectiveness of this approach is justified by the word clouds presented in Figures 7 and 8.

2. The prior works [59,60] on sentiment analysis in the context of tweets about COVID-19 focused on developing approaches for classifying the sentiment only into three classes - 'positive', 'negative', and 'neutral'. In a realistic scenario, there can be different kinds of 'positive' emotions, such as 'good' and 'great'. Similarly, there can be different kinds of 'negative' emotions, such as 'bad' and 'terrible'. The existing works cannot differentiate between these kinds of positive (or negative) emotions. To address the need in this context, associated with increasing the number of classes for classification of the sentiment of the tweet, this work proposes an approach that classifies tweets into five sentiment classes: 'great', 'good', 'neutral', 'bad' and 'terrible.'
3. The emerging works in this field, for instance, related to detecting trending topics [61], anomaly events [62], public perceptions towards C.D.C. [63], and views towards not wearing masks [64], focused on the development of new frameworks and methodologies without focusing on quantifying multimodal components of the characteristics of the tweets and ranking these characteristics to infer insights about social media activity on Twitter due to COVID-19. This work addresses this need. The results from sentiment analysis, type detection, source tracking, language interpretation, and embedded URLs observation of the tweets were categorized into distinct categories, and these categories were ranked in terms of the associated characteristics to infer meaningful and relevant insights about social media activity on Twitter related to tweets posted about the SARS-CoV-2 Omicron variant. For instance, in the tweet type analysis, the findings of this study show that "Twitter for Android" accounted for the most number of tweets (35.2% of the total tweets), which was followed by "Twitter Web App" (29.2% of the total tweets), "Twitter for iPhone" (25.8% of the total tweets), and other sources.
4. The previous works centered around performing data analysis on tweets related to COVID-19 included a small corpus of tweets, such as 4081 tweets in [73] and 4492 tweets in [71]. In view of the number of active Twitter users and the number of tweets posted each day, there is a need to include more tweets in the data analysis process. This work addresses this need by considering a total of 12028 relevant tweets that had a combined reach of 149,500,959, 226,603,833 impressions, 1,053,869 retweets, and 3,427,976 favorites.

The above results present several characteristics, insights, and features regarding tweets about the COVID-19 omicron variant. These findings are expected to have a wide range of applications and provide evidence for the investigation of multiple research questions in the areas of Natural Language Processing, Big Data, Data Mining, Data Science, Machine Learning, Artificial Intelligence, and their related areas with a specific focus on studying, analyzing, investigating, and interpreting social media activity, behavior, communications, and conversations in the context of information seeking, sharing, and exchange related to the SARS-CoV-2 omicron variant.

5. Conclusions

Since the initial outbreak in Wuhan, China, in December 2019, the SARS-CoV-2 virus has resulted in a total of 520,628,068 cases and 6,287,207 deaths on a global scale. The virus has undergone multiple mutations, and as a result, several variants have been detected, such as Alpha, Beta, Gamma, Delta, Epsilon, Eta, Iota, Kappa, Zeta, Mu, and Omicron, in different parts of the world. Out of these variants, the Omicron variant, a variant of concern (VOC) as per WHO, is currently the globally dominant variant and has been reported to be the most immune evasive VOC of SARS-CoV-2 and is also considered to be highly resistant against antibody-mediated neutralization. The number of cases and deaths due to Omicron in different parts of the world is on a constant rise.

Research conducted during pandemics in the past suggests that people extensively use social media platforms for communication, information, news, views, opinions, ideas, knowledge, feedback, and experiences related to the pandemic they are facing. Twitter, one such social media platform, is popular amongst all age groups. Therefore, this work took a comprehensive approach to identify, study, and analyze tweets related to the SARS-CoV-2 Omicron variant to understand, categorize, and interpret the associated dynamics and characteristic features of social media behavior. A total of 12028 tweets about the SARS-CoV-2 Omicron variant were mined from Twitter, and the associated sentiment ('great', 'good', 'neutral', 'bad', and 'terrible'), type (original tweet, retweet, or reply), source (such as "Twitter for Android", "Twitter Web App", "Twitter for iPhone", etc.), language (such as English, Spanish, Bengali, etc.), and embedded URLs (URLs that were included in the tweet text) were analyzed. The findings from this exploratory study are manifold. First, a majority of the tweets had a 'neutral' emotion (50.5% of the total tweets), which was followed by 'bad' (15.6% of the total tweets), 'good' (14.0% of the total tweets), 'terrible' (12.5% of the total tweets), and 'great' (7.5% of the total tweets). Second, 35.2% of the tweets had "Twitter for Android" as their source. It was followed by the "Twitter Web App" (29.2% of the total tweets), "Twitter for iPhone" (25.8% of the total tweets), "Twitter for iPad" (3.8% of the total tweets), "Tweetdeck" (1.6% of the total tweets), and other sources which accounted for less than 1% of the total tweets. Third, a majority of the tweets (65.9%) were posted in English, which was followed by Spanish or Castillian (10.5% of the total tweets), French (5.1% of the total tweets), Italian (3.3% of the total tweets), Japanese (2.5% of the total tweets), and other languages that accounted for less than 2% of the tweets. Fourth, the majority of the tweets (60.8%) in terms of tweet type were retweets which were followed by original tweets (19.8% of the total tweets) and replies (19.4% of the total tweets). Finally, in terms of embedded URLs, the most common URL embedded in the tweets was found to be twitter.com, which was followed by biorxiv.org, nature.com, wapo.st, nzherald.co.nz, recvprofits.com, science.org, and a few other sources. In addition to the above findings, the work also addresses the limitations in the prior works related to studying social media behavior on Twitter in the context of COVID-19.

As the SARS-CoV-2 omicron variant continues to spread across the globe, rapid advances are being made related to omicron-specific vaccines [89,90], and new studies related to this variant are also getting published [91, 92]. These advances, studies, new findings, and the nature of public reaction, views, opinion, feedback, thoughts, and perspectives towards the same may impact one or more of the characteristic features of social media behavior on Twitter that were investigated and analyzed in this study. To address this limitation, a follow-up study will be conducted in the future. In that follow-up study, the aim would be to gather all the relevant tweets in between the dates when the first case of Omicron was recorded and when the last case of Omicron would be recorded to perform this exploratory analysis once again to compare the associated findings and to comment on any similarities and dissimilarities in the insights that might be observed.

Author Contributions: Conceptualization, NT; methodology, N.T.; software, N.T.; validation, N.T.; formal analysis, N.T.; investigation, N.T.; resources, N.T.; data curation, N.T.; visualization, N.T.; data analysis and results, N.T.; writing—original draft preparation, N.T.; writing—review and editing, N.T.; supervision, C.Y.H.; project administration, C.Y.H.; funding acquisition, Not Applicable. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ksiazek, T.G.; Erdman, D.; Goldsmith, C.S.; Zaki, S.R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J.A.; Lim, W.; et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. N. Engl. J. Med. 2003, 348, 1953–1966, doi:10.1056/NEJMoa030781.
2. Fauci, A.S.; Lane, H.C.; Redfield, R.R. Covid-19 - Navigating the Uncharted. N. Engl. J. Med. 2020, 382, 1268–1269, doi:10.1056/NEJMe2002387.
3. Wilder-Smith, A.; Chiew, C.J.; Lee, V.J. Can We Contain the COVID-19 Outbreak with the Same Measures as for SARS? Lancet Infect. Dis. 2020, 20, e102–e107, doi:10.1016/S1473-3099(20)30129-8.
4. Cucinotta, D.; Vanelli, M. WHO Declares COVID-19 a Pandemic. Acta Biomed. 2020, 91, 157–160, doi:10.23750/abm.v91i1.9397.
5. COVID Live - Coronavirus Statistics - Worldometer Available online: https://www.worldometers.info/coronavirus/ (accessed on 14 May 2022).
6. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. Lancet 2020, 395, 497–506, doi:10.1016/S0140-6736(20)30183-5.
7. Lauring, A.S.; Hodcroft, E.B. Genetic Variants of SARS-CoV-2-What Do They Mean? JAMA 2021, 325, 529–531, doi:10.1001/jama.2020.27124.
8. Khare, S.; Gurry, C.; Freitas, L.; Schultz, M.B.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; Lee, R.T.; Yeo, W.; et al. GISAID's Role in Pandemic Response. China CDC Wkly 2021, 3, 1049–1051, doi:10.46234/ccdcw2021.255.
9. GISAID - Initiative Available online: https://www.gisaid.org/ (accessed on 14 May 2022).
10. C.D.C. SARS-CoV-2 Variant Classifications and Definitions Available online: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html (accessed on 14 May 2022).
11. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern Available online: https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern (accessed on 14 May 2022).
12. Gobeil, S.M.-C.; Henderson, R.; Stalls, V.; Janowska, K.; Huang, X.; May, A.; Speakman, M.; Beaudoin, E.; Manne, K.; Li, D.; et al. Structural Diversity of the SARS-CoV-2 Omicron Spike. Mol. Cell 2022, doi:10.1016/j.molcel.2022.03.028.
13. Schmidt, F.; Muecksch, F.; Weisblum, Y.; Da Silva, J.; Bednarski, E.; Cho, A.; Wang, Z.; Gaebler, C.; Caskey, M.; Nussenzweig, M.C.; et al. Plasma Neutralization of the SARS-CoV-2 Omicron Variant. N. Engl. J. Med. 2022, 386, 599–601, doi:10.1056/NEJMc2119641.
14. Hoffmann, M.; Krüger, N.; Schulz, S.; Cossmann, A.; Rocha, C.; Kempf, A.; Nehlmeier, I.; Graichen, L.; Moldenhauer, A.-S.; Winkler, M.S.; et al. The Omicron Variant Is Highly Resistant against Antibody-Mediated Neutralization: Implications for Control of the COVID-19 Pandemic. Cell 2022, 185, 447-456.e11, doi:10.1016/j.cell.2021.12.032.
15. Francis, A.I.; Ghany, S.; Gilkes, T.; Umakanthan, S. Review of COVID-19 Vaccine Subtypes, Efficacy and Geographical Distributions. Postgrad. Med. J. 2022, 98, 389–394, doi:10.1136/postgradmedj-2021-140654.
16. Gavriatopoulou, M.; Ntanasis-Stathopoulos, I.; Korompoki, E.; Fotiou, D.; Migkou, M.; Tzanninis, I.-G.; Psaltopoulou, T.; Kastritis, E.; Terpos, E.; Dimopoulos, M.A. Emerging Treatment Strategies for COVID-19 Infection. Clin. Exp. Med. 2021, 21, 167–179, doi:10.1007/s10238-020-00671-y.
17. Weekly Epidemiological Update on COVID-19 - 22 March 2022 Available online: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---22-march-2022 (accessed on 14 May 2022).
18. Feiner, L. WHO Says Omicron Cases Are "off the Charts" as Global Infections Set New Records Available online: https://www.cnbc.com/2022/01/12/who-says-omicron-cases-are-off-the-charts-as-global-infections-set-new-records.html (accessed on 14 May 2022).
19. SARS-CoV-2 Omicron Variant Cases Worldwide 2022 Available online: https://www.statista.com/statistics/1279100/number-omicron-variant-worldwide-by-country/ (accessed on 14 May 2022).
20. Katz, M.; Nandi, N. Social Media and Medical Education in the Context of the COVID-19 Pandemic: Scoping Review. JMIR Med. Educ. 2021, 7, e25892, doi:10.2196/25892.
21. Sharma, M.; Yadav, K.; Yadav, N.; Ferdinand, K.C. Zika Virus Pandemic—Analysis of Facebook as a Social Media Health Information Platform. Am. J. Infect. Control 2017, 45, 301–302, doi:10.1016/j.ajic.2016.08.022.
22. Wiederhold, B.K. Social Media and Social Organizing: From Pandemic to Protests. Cyberpsychol. Behav. Soc. Netw. 2020, 23, 579–580, doi:10.1089/cyber.2020.0461.
23. Fung, I.C.-H.; Duke, C.H.; Finch, K.C.; Snook, K.R.; Tseng, P.-L.; Hernandez, A.C.; Gambhir, M.; Fu, K.-W.; Tse, Z.T.H. Ebola Virus Disease and Social Media: A Systematic Review. Am. J. Infect. Control 2016, 44, 1660–1671, doi:10.1016/j.ajic.2016.05.011.
24. Ding, H. Social Media and Participatory Risk Communication during the H1N1 Flu Epidemic: A Comparative Study of the United States and China. China Media Rep. Overseas 2010, 6, 80+.
25. Longley, P.A.; Adnan, M.; Lansley, G. The Geotemporal Demographics of Twitter Usage. Environ. Plan. A 2015, 47, 465–484, doi:10.1068/a130122p.
26. The Latest Twitter Statistics: Everything You Need to Know — DataReportal – Global Digital Insights Available online: https://datareportal.com/essential-twitter-stats (accessed on 14 May 2022).
27. Lazard, A.J.; Scheinfeld, E.; Bernhardt, J.M.; Wilcox, G.B.; Suran, M. Detecting Themes of Public Concern: A Text Mining Analysis of the Centers for Disease Control and Prevention's Ebola Live Twitter Chat. Am. J. Infect. Control 2015, 43, 1109–1111, doi:10.1016/j.ajic.2015.05.025.
28. Bolotova, Y.V.; Lou, J.; Safro, I. Detecting and Monitoring Foodborne Illness Outbreaks: Twitter Communications and the 2015 U.S. Salmonella Outbreak Linked to Imported Cucumbers. arXiv [stat. A.P.] 2017, doi:10.48550/ARXIV.1708.07534.

29. Gomide, J.; Veloso, A.; Meira, W., Jr; Almeida, V.; Benevenuto, F.; Ferraz, F.; Teixeira, M. Dengue Surveillance Based on a Computational Model of Spatio-Temporal Locality of Twitter. In Proceedings of the Proceedings of the 3rd International Web Science Conference on - WebSci '11; A.C.M. Press: New York, New York, U.S.A., 2011.
30. Tomaszewski, T.; Morales, A.; Lourentzou, I.; Caskey, R.; Liu, B.; Schwartz, A.; Chin, J. Identifying False Human Papillomavirus (HPV) Vaccine Information and Corresponding Risk Perceptions from Twitter: Advanced Predictive Models. J. Med. Internet Res. 2021, 23, e30451, doi:10.2196/30451.
31. Do, H.J.; Lim, C.-G.; Kim, Y.J.; Choi, H.-J. Analyzing Emotions in Twitter during a Crisis: A Case Study of the 2015 Middle East Respiratory Syndrome Outbreak in Korea. In Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp); IEEE, 2016; pp. 415–418.
32. Radzikowski, J.; Stefanidis, A.; Jacobsen, K.H.; Croitoru, A.; Crooks, A.; Delamater, P.L. The Measles Vaccination Narrative in Twitter: A Quantitative Analysis. JMIR Public Health Surveill. 2016, 2, e1, doi:10.2196/publichealth.5059.
33. Fu, K.-W.; Liang, H.; Saroha, N.; Tse, Z.T.H.; Ip, P.; Fung, I.C.-H. How People React to Zika Virus Outbreaks on Twitter? A Computational Content Analysis. Am. J. Infect. Control 2016, 44, 1700–1702, doi:10.1016/j.ajic.2016.04.253.
34. Signorini, A.; Segre, A.M.; Polgreen, P.M. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. PLoS One 2011, 6, e19467, doi:10.1371/journal.pone.0019467.
35. Gesualdo, F.; Stilo, G.; Agricola, E.; Gonfiantini, M.V.; Pandolfi, E.; Velardi, P.; Tozzi, A.E. Influenza-like Illness Surveillance on Twitter through Automated Learning of Naïve Language. PLoS One 2013, 8, e82489, doi:10.1371/journal.pone.0082489.
36. Szomszor, M.; Kostkova, P.; de Quincey, E. #swineflu: Twitter Predicts Swine Flu Outbreak in 2009. In Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; pp. 18–26 ISBN 9783642236341.
37. Alessa, A.; Faezipour, M. Flu Outbreak Prediction Using Twitter Posts Classification and Linear Regression with Historical Centers for Disease Control and Prevention Reports: Prediction Framework Study. JMIR Public Health Surveill. 2019, 5, e12383, doi:10.2196/12383.
38. Hirschfeld, D. Twitter Data Accurately Tracked Haiti Cholera Outbreak. Nature 2012, doi:10.1038/nature.2012.9770.
39. Van Der Vyver, A.G. The Listeriosis Outbreak in South Africa: A Twitter Analysis of Public Reaction Available online: http://www.icmis.net/icmis18/ICMIS18CD/pdf/S198-final.pdf (accessed on 14 May 2022).
40. Thackeray, R.; Burton, S.H.; Giraud-Carrier, C.; Rollins, S.; Draper, C.R. Using Twitter for Breast Cancer Prevention: An Analysis of Breast Cancer Awareness Month. BMC Cancer 2013, 13, 508, doi:10.1186/1471-2407-13-508.
41. Da, B.L.; Surana, P.; Schueler, S.A.; Jalaly, N.Y.; Kamal, N.; Taneja, S.; Vittal, A.; Gilman, C.L.; Heller, T.; Koh, C. Twitter as a Noninvasive Bio-Marker for Trends in Liver Disease. Hepatol. Commun. 2019, 3, 1271–1280, doi:10.1002/hep4.1394.
42. Khan, A.; Silverman, A.; Rowe, A.; Rowe, S.; Tick, M.; Testa, S.; Dodds, K.; Alabbas, B.; Borum, M.L. Who Is Saying What about Inflammatory Bowel Disease on Twitter? In Proceedings of the G.W. Research Days 2016 - 2020; 2018.
43. McLean, R.; Shirazian, S. Women and Kidney Disease: A Twitter Conversation for One and All. Kidney Int. Rep. 2018, 3, 767–768, doi:10.1016/j.ekir.2018.05.001.
44. Stens, O.; Weisman, M.H.; Simard, J.; Reuter, K. Insights from Twitter Conversations on Lupus and Reproductive Health: Protocol for a Content Analysis. JMIR Res. Protoc. 2020, 9, e15623, doi:10.2196/15623.
45. Cevik, F.; Kilimci, Z.H. Analysis of Parkinson's Disease Using Deep Learning and Word Embedding Models. acperpro 2019, 2, 786–797, doi:10.33793/acperpro.02.03.86.
46. Porat, T.; Garaizar, P.; Ferrero, M.; Jones, H.; Ashworth, M.; Vadillo, M.A. Content and Source Analysis of Popular Tweets Following a Recent Case of Diphtheria in Spain. Eur. J. Public Health 2019, 29, 117–122, doi:10.1093/eurpub/cky144.
47. Sugumaran, R.; Voss, J. Real-Time Spatio-Temporal Analysis of West Nile Virus Using Twitter Data. In Proceedings of the Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications - C.O.M.Geo '12; A.C.M. Press: New York, New York, U.S.A., 2012.
48. Rahmanian, V.; Jahanbin, K. Using Twitter and Web News Mining to Predict COVID-19 Outbreak. Asian Pac. J. Trop. Med. 2020, 13, 378, doi:10.4103/1995-7645.279651.
49. Rosenberg, H.; Syed, S.; Rezaie, S. The Twitter Pandemic: The Critical Role of Twitter in the Dissemination of Medical Information and Misinformation during the COVID-19 Pandemic. CJEM 2020, 22, 418–421, doi:10.1017/cem.2020.361.
50. Haman, M. The Use of Twitter by State Leaders and Its Impact on the Public during the COVID-19 Pandemic. Heliyon 2020, 6, e05540, doi:10.1016/j.heliyon.2020.e05540.
51. Alhayan, F.; Pennington, D.; Ayouni, S. Twitter Use by the Dementia Community during COVID-19: A User Classification and Social Network Analysis. Online Inf. Rev. 2022, ahead-of-print, 88, doi:10.1108/oir-04-2021-0208.
52. Guo, J.-W.; Radloff, C.L.; Wawrzynski, S.E.; Cloyes, K.G. Mining Twitter to Explore the Emergence of COVID-19 Symptoms. Public Health Nurs. 2020, 37, 934–940, doi:10.1111/phn.12809.
53. Roy, S.; Ghosh, P. A Comparative Study on Distancing, Mask and Vaccine Adoption Rates from Global Twitter Trends. Healthcare (Basel) 2021, 9, 488, doi:10.3390/healthcare9050488.
54. Yousefinaghani, S.; Dara, R.; Mubareka, S.; Papadopoulos, A.; Sharif, S. An Analysis of COVID-19 Vaccine Sentiments and Opinions on Twitter. Int. J. Infect. Dis. 2021, 108, 256–262, doi:10.1016/j.ijid.2021.05.059.
55. Wang, S.; Schraagen, M.; Sang, E.T.K.; Dastani, M. Dutch General Public Reaction on Governmental COVID-19 Measures and Announcements in Twitter Data. arXiv [cs.SI] 2020, doi:10.48550/ARXIV.2006.07283.
56. Krittanawong, C.; Narasimhan, B.; Virk, H.U.H.; Narasimhan, H.; Hahn, J.; Wang, Z.; Tang, W.H.W. Misinformation Dissemination in Twitter in the COVID-19 Era. Am. J. Med. 2020, 133, 1367–1369, doi:10.1016/j.amjmed.2020.07.012.

57. Ahmed, W.; Vidal-Alaball, J.; Downing, J.; López Seguí, F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J. Med. Internet Res. 2020, 22, e19458, doi:10.2196/19458.
58. Bonnevie, E.; Gallegos-Jeffrey, A.; Goldbarg, J.; Byrd, B.; Smyser, J. Quantifying the Rise of Vaccine Opposition on Twitter during the COVID-19 Pandemic. J. Commun. Healthc. 2021, 14, 12–19, doi:10.1080/17538068.2020.1858222.
59. Mehedi Shamrat, F.M.J.; Chakraborty, S.; Imran, M.M.; Muna, J.N.; Billah, M.M.; Das, P.; Rahman, M.O. Sentiment Analysis on Twitter Tweets about COVID-19 Vaccines Usi Ng N.L.P. and Supervised K.N.N. Classification Algorithm. Indones. j. electr. eng. comput. sci. 2021, 23, 463, doi:10.11591/ijeecs.v23.i1.pp463-470.
60. Sontayasara, T.; Jariyapongpaiboon, S.; Promjun, A.; Seelpipat, N.; Saengtabtim, K.; Tang, J.; Leelawat, N.; Department of Industrial Engineering, Faculty of Engineering, Chulalongkorn University 254 Phayathai Road, Pathumwan, Bangkok 10330, Thailand; International School of Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand; Disaster and Risk Management Information Systems Research Group, Chulalongkorn University, Bangkok, Thailand Twitter Sentiment Analysis of Bangkok Tourism during COVID-19 Pandemic Using Support Vector Machine Algorithm. J. Disaster Res. 2021, 16, 24–30, doi:10.20965/jdr.2021.p0024.
61. Asgari-Chenaghlu, M.; Nikzad-Khasmakhi, N.; Minaee, S. Covid-Transformer: Detecting COVID-19 Trending Topics on Twitter Using Universal Sentence Encoder. arXiv [cs.CL] 2020, doi:10.48550/ARXIV.2009.03947.
62. Amen, B.; Faiz, S.; Do, T.-T. Big Data Directed Acyclic Graph Model for Real-Time COVID-19 Twitter Stream Detection. Pattern Recognit. 2022, 123, 108404, doi:10.1016/j.patcog.2021.108404.
63. Lyu, J.C.; Luli, G.K. Understanding the Public Discussion about the Centers for Disease Control and Prevention during the COVID-19 Pandemic Using Twitter Data: Text Mining Analysis Study. J. Med. Internet Res. 2021, 23, e25108, doi:10.2196/25108.
64. Al-Ramahi, M.; Elnoshokaty, A.; El-Gayar, O.; Nasralah, T.; Wahbeh, A. Public Discourse against Masks in the COVID-19 Era: Infodemiology Study of Twitter Data. JMIR Public Health Surveill. 2021, 7, e26780, doi:10.2196/26780.
65. Jain, S.; Sinha, A. Identification of Influential Users on Twitter: A Novel Weighted Correlated Influence Measure for Covid-19. Chaos Solitons Fractals 2020, 139, 110037, doi:10.1016/j.chaos.2020.110037.
66. Madani, Y.; Erritali, M.; Bouikhalene, B. Using Artificial Intelligence Techniques for Detecting Covid-19 Epidemic Fake News in Moroccan Tweets. Results Phys. 2021, 25, 104266, doi:10.1016/j.rinp.2021.104266.
67. Shokoohyar, S.; Rikhtehgar Berenji, H.; Dang, J. Exploring the Heated Debate over Reopening for Economy or Continuing Lockdown for Public Health Safety Concerns about COVID-19 in Twitter. Int. J. Bus. Syst. Res. 2021, 15, 650, doi:10.1504/ijbsr.2021.117316.
68. Chehal, D.; Gupta, P.; Gulati, P. COVID-19 Pandemic Lockdown: An Emotional Health Perspective of Indians on Twitter. Int. J. Soc. Psychiatry 2021, 67, 64–72, doi:10.1177/0020764020940741.
69. Glowacki, E.M.; Wilcox, G.B.; Glowacki, J.B. Identifying #addiction Concerns on Twitter during the COVID-19 Pandemic: A Text Mining Analysis. Subst. Abus. 2021, 42, 39–46, doi:10.1080/08897077.2020.1822489.
70. Selman, L.E.; Chamberlain, C.; Sowden, R.; Chao, D.; Selman, D.; Taubert, M.; Braude, P. Sadness, Despair and Anger When a Patient Dies Alone from COVID-19: A Thematic Content Analysis of Twitter Data from Bereaved Family Members and Friends. Palliat. Med. 2021, 35, 1267–1276, doi:10.1177/02692163211017026.
71. Koh, J.X.; Liew, T.M. How Loneliness Is Talked about in Social Media during COVID-19 Pandemic: Text Mining of 4,492 Twitter Feeds. J. Psychiatr. Res. 2022, 145, 317–324, doi:10.1016/j.jpsychires.2020.11.015.
72. Mackey, T.; Purushothaman, V.; Li, J.; Shah, N.; Nali, M.; Bardier, C.; Liang, B.; Cai, M.; Cuomo, R. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated with COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health Surveill. 2020, 6, e19509, doi:10.2196/19509.
73. Leung, J.; Chung, J.Y.C.; Tisdale, C.; Chiu, V.; Lim, C.C.W.; Chan, G. Anxiety and Panic Buying Behaviour during COVID-19 Pandemic-A Qualitative Analysis of Toilet Paper Hoarding Contents on Twitter. Int. J. Environ. Res. Public Health 2021, 18, 1127, doi:10.3390/ijerph18031127.
74. Banda, J.M.; Tekumalla, R.; Wang, G.; Yu, J.; Liu, T.; Ding, Y.; Chowell, G. A Large-Scale COVID-19 Twitter Chatter Dataset for Open Scientific Research -- an International Collaboration. ArXiv 2020, 2, 315–324, doi:10.3390/epidemiologia2030024.
75. 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.
76. Elhadad, M.K.; Li, K.F.; Gebali, F. COVID-19-FAKES: A Twitter (Arabic/English) Dataset for Detecting Misleading Information on COVID-19. In Advances in Intelligent Networking and Collaborative Systems; Springer International Publishing: Cham, 2021; pp. 256–268 ISBN 9783030577957.
77. Hayawi, K.; Shahriar, S.; Serhani, M.A.; Taleb, I.; Mathew, S.S. ANTi-Vax: A Novel Twitter Dataset for COVID-19 Vaccine Misinformation Detection. Public Health 2022, 203, 23–30, doi:10.1016/j.puhe.2021.11.022.
78. Privacy Policy of Twitter Available online: https://twitter.com/en/privacy/previous/version_15 (accessed on 15 May 2022).
79. Twitter Developer Agreement and Policy Available online: https://developer.twitter.com/en/developer-terms/agreement-and-policy (accessed on 15 May 2022).
80. Standard Search API Available online: https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets (accessed on 15 May 2022).
81. How to Tweet Available online: https://help.twitter.com/en/using-twitter/how-to-tweet (accessed on 15 May 2022).
82. Twitter Official Website Available online: https://twitter.com/ (accessed on 15 May 2022).
83. Twitter Android Application Available online: https://play.google.com/store/apps/details?id=com.twitter.android&hl=en_US&gl=US (accessed on 15 May 2022).

84. Twitter for IPhone Available online: https://apps.apple.com/in/app/twitter/id333903271 (accessed on 15 May 2022).
85. Wikipedia contributors TweetDeck Available online: https://en.wikipedia.org/w/index.php?title=TweetDeck&oldid=1056092943.
86. Supported Languages on Twitter Available online: https://developer.twitter.com/en/docs/twitter-for-websites/supported-languages (accessed on 15 May 2022).
87. Social Bearing Research Tool Available online: https://socialbearing.com/ (accessed on 15 May 2022).
88. Vijay Karunamurthy Make Way for Youtu.Be Links Available online: https://blog.youtube/news-and-events/make-way-for-youtube-links/ (accessed on 15 May 2022).
89. Moderna Begins next Phase of Omicron-Specific Booster Trial as Study Finds That Antibodies Remain Durable despite 6-Fold Drop over 6 Months Available online: https://www.cnn.com/2022/01/26/health/moderna-omicron-antibodies-booster/index.html (accessed on 16 May 2022).
90. Pfizer and BioNTech Initiate Study to Evaluate Omicron-Based COVID-19 Vaccine in Adults 18 to 55 Years of Age Available online: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-initiate-study-evaluate-omicron-based (accessed on 16 May 2022).
91. Strasser, Z.; Hadavand, A.; Murphy, S.; Estiri, H. SARS-CoV-2 Omicron Variant Is as Deadly as Previous Waves after Adjusting for Vaccinations, Demographics, and Comorbidities. Research Square 2022.
92. Bar-On, Y.M.; Goldberg, Y.; Mandel, M.; Bodenheimer, O.; Amir, O.; Freedman, L.; Alroy-Preis, S.; Ash, N.; Huppert, A.; Milo, R. Protection by a Fourth Dose of BNT162b2 against Omicron in Israel. N. Engl. J. Med. 2022, 386, 1712–1720, doi:10.1056/NEJMoa2201570.