Sentiment analysis of Indian Tweets about Covid-19 vaccines

Aasif Ahmad Mir
Department of Library and Information Science, Pondicherry University, India

Rathinam Sevukan
Department of Library and Information Science, Pondicherry University, India

Abstract

People are becoming more reliant on social media networks to express their opinions about various topics and obtain health information. The study is intended to explore and analyse the sentiments of Indian people related to Covid-19 vaccines as well as to visualise the top most frequently occurring terms individuals have used to communicate their ideas on Twitter about Covid-19 vaccines in India. The Tweet Archiver was used to retrieve the Tweets against 'Covid19vaccine' and 'Coronavirusvaccine' hashtags for the period of 2 months 18 days (4 January 2021–22 March 2021). After collecting data, the Orange software and VOSviewer were used for further analysis. The Tweets were posted across the country, with an immense contribution from Maharashtra (223, 15.58%), followed by Delhi (220, 15.37%) and Tamil Nadu (73, 5.10%). The majority (639, 44.65%) of the Tweets reflect positive sentiments, followed by neutral (521, 38.50%) and negative (241, 16.84%) sentiments, respectively. This signifies that most Twitter users have a favourable opinion towards Covid vaccines in India. Based on the relevance score of the words, the words 'Delhi heart', 'Lung institute', 'Gift', 'Unite2fightcorona', and 'Covid-19 Vaccine' are the leading words appearing in Tweets. The study illustrates the sentiments of the Indian people towards 'Covid-19 vaccines', gains some insights into overall public communication about the topic and complements the existing literature. It can assist health policymakers and administrators in better understanding the polarity (positive, negative, and neutral) of Tweets about Covid-19 vaccines on Twitter to raise public awareness about health concerns and misinformation about the vaccine.

Keywords

Covid-19 vaccine; health informatics; micro-blogging; public opinions; sentiment analysis; social media; Twitter

1. Introduction

Coronaviruses are opportunistic infections that affect people with weak immune systems. It was first identified in the 1960s [1]. Coronaviruses are enveloped, non-segmented positive-sense RNA virus belonging to the order *Nidovirales*, family *Coronaviridae* and the subfamily *Coronavirinae* [1–3]. It is widely distributed among humans, other mammals and birds and causes respiratory, enteric, hepatic and neurological diseases [4,5]. Six coronavirus species are known to afflict humans [6]. The 2019 coronavirus disease (Covid-19), a contagious disease that poses a serious threat to the global population, was discussed worldwide [7–10]. Its outbreak began with a cluster of pneumonia cases in Wuhan, China, caused by a novel betacoronavirus, the 2019 novel coronavirus (2019-nCoV) [11]. The patients with pneumonia of unknown cause were reported that were epidemiologically connected to the seafood and wet animal wholesale market in the Central Chinese province of Hubei in December 2019 [12–14]. Since then, it has spread to several other global locations. On 12 January 2020, the World Health Organization (WHO) temporarily termed the new virus '2019 novel coronavirus' (2019-nCoV) and then, on 12 February 2020, officially named this infectious disease 'coronavirus disease 2019' (Covid-19). After analysing and reviewing the situation, WHO, on 11 March 2020, announced the Covid-19 outburst as a pandemic [15]. Later on, the 2019-nCoV name was changed to 'Severe Acute Respiratory Syndrome Coronavirus-2

Corresponding author:
Aasif Ahmad Mir, Department of Library and Information Science, Pondicherry University, Kalapet 605 014, Puducherry, India.
Email: miraasif7298@gmail.com

(SARS-CoV-2) by the International Committee on Taxonomy of Viruses' [16]. The Covid-19 pandemic has dramatically affected and influenced the masses across the globe [17]. The health care initiatives have been planned at a global level. Therefore, it will not be wrong to say that Covid-19 has had an unparalleled impact on global society [17,18].

'The global rollout of Covid-19 vaccines sparked feelings of relief and newfound optimism for so many. The discussion of vaccination progress, accessibility, efficacy, and side effects was ongoing, permeating news stories and Twitter spheres every day' [19]. Social media has evolved into an essential platform for individuals, companies and governments worldwide to interact and express themselves on different topics [20]. The Covid-19 pandemic has prompted an increase in the use of social media to address a variety of pandemic-related issues, including vaccinations [21]. Twitter has been chosen as the primary source due to its prominence as a platform for health-related discussion and information [15,22,23] and for being a popular social media platform for academics and scientists to investigate various perspectives on human views [24] and is well-known as a backchannel activity platform [25]. 'Twitter, currently the leading microblogging social network, has attracted a great body of research works' [26] and has become a popular subject of research that has been widely investigated in the literary world in various disciplines ranging from the Social Sciences to the Health Sciences [27]. It has become a supportive online media and has evolved into a cross-border information transporter, allowing stakeholders and administrators to respond quickly to a crisis [28]. It has become an essential data source for spotting various outbreaks and understanding of public attitudes and behaviour during emergencies [29,30].

Sentiment Analysis, also known as Opinion Mining, is a computer study of people's views, attitudes and emotions that may be used to display public impressions of any entity. During the pandemic, people used Twitter to express their opinions about Covid-19 and its vaccines [31]. So, the Twitter data were utilised in this study to investigate the opinions of Indian people about Covid-19 vaccines. While it is not necessary that Twitter can represent the general public's opinion, it can represent the many forms of news that people are exposed to and serve as a key source of discussion regarding the Covid-19 vaccines [22,32].

2. Review of literature

Infectious disease outbreaks and other health problems were widely discussed and shared on social media [33,34]. Sentiment analysis of social media data is emerging as an important research field and is in use at different spheres, namely, e-sports [35,36], sports [37–41] healthcare [42–44] and even geopolitical conflicts and crises [45–50]. Since the Covid-19 outbreak, a number of studies throughout the world have utilised Twitter data to assess the public opinion [21,28,51–57], main discussion themes [58–61] and misinformation [62–64] about the Covid-19 vaccine. Marcec and Likic [51] analysed sentiments regarding various Covid-19 vaccines employing AFINN lexicon strategy to find out the opinion of masses regarding various vaccines that were developed over a period of time for Covid-19 and concluded that people showed positive sentiments towards Pfizer and Moderna vaccines while negative sentiments developed for AstraZeneca/Oxford vaccine over time. Shamrat et al. [31] examined the public perception of Covid-19 vaccines developed by Pfizer, Moderna and AstraZeneca by utilising NLP and supervised KNN algorithms. They revealed that people are more optimistic about Pfizer and Moderna vaccines than AstraZeneca vaccines. Trisnawati et al. [65] investigated the usage of the Sinovac vaccine among Indonesian people using the Naive Bayes approach. They discovered that, in general, the view of the masses about the vaccination was primarily neutral and positive rather than having a negative attitude about the Sinovac vaccine. Aryal and Bhattarai [66] investigated the attitudes of individuals on Twitter regarding Covid-19 vaccination. They found that majority of people, especially from Nepal, India and Singapore, showed the most positive sentiments towards Covid-19 vaccination than the people from other Asian countries. Dubey [67] examined the opinion of people in India about Oxford-AstraZeneca's Covishield and Bharat Biotech's Covaxin vaccines that were developed as a protective measure against Covid-19. He revealed that people showed more positive sentiments and trust for Covishield than Covaxin. Alselwi and Kaynak [68] carried out a sentiment analysis of Twitter data on Covid-19 vaccines using Natural Language Processing. They found that many people exhibited neutral sentiments while a meagre number of people had negative opinions than positive ones. Ansari and Khan [54] analysed people's opinions on Covid-19 vaccines around the globe. They found that most countries had positive sentiments towards vaccination drives while a few countries had both positive and negative opinions regarding Covid-19 vaccines. They also found that the attitude of males towards the different vaccines was more positive than that of females. Hussain et al. [69] studied the attitudes of Tweets posted by users of the United States and United Kingdom regarding Covid-19 vaccines, mainly focusing on utilising the amalgamation of Natural Language Processing with Machine learning approach, and inferred that people of the United Kingdom had greater positive sentiments about the vaccines than the masses in the United States. Hu et al. [56] also investigated the opinion of Twitter users of the United States on Covid-19 vaccines and concluded that throughout the time, people largely had positive sentiments, from most of the states, about the vaccines developed for Covid-19 and that negative sentiments among individuals could be seen decreasing at the same time. Mahanti et al. [70] while doing a

comparative sentiment analysis of the Tweets among five countries – India, Canada, Australia, the United States, and the United Kingdom, discovered that Canada was the only country that had unfavourable opinions about Covid-19 vaccinations over the study period. In contrast, other countries showed positive sentiments, particularly India and Australia, towards vaccination. Kwok et al. [71] also analysed the sentiments of Australians about Covid-19 vaccines. They discovered that most people have negative opinions about vaccines. The study attributed that such people may not know the severity of the life-threatening disease, that is, Covid-19, so they were more inclined to support and spread negative emotions for Covid-19 vaccines. Gulati [72] investigated vaccine tourism on Twitter. He noted that the maximum of the Tweets carried positive sentiments, and the least had negative opinions regarding vaccine tourism, suggesting overall positive sentiments on Twitter. Reviewing the earlier studies on the Covid-19 vaccinations revealed that no specific study had focused on the views of individual Indian states regarding the Covid-19 vaccines. Therefore, the present study is indispensable.

3. Objectives

The objectives of this study are to

1. analyse sentiments expressed by Indian people about Covid-19 vaccines and
2. visualise the representation of text appearing in Tweets.

4. Methodology

The approach used to complete the study is divided into five phases.

4.1. Phase I: hashtag selection for tweet delimitation

The two random hashtags '#Covid19Vaccine' and '#CoronaVirusVaccine' were used as search keywords to download Tweets.

4.2. Phase II: harvesting of Tweets

The 'Tweet Archiver' was used to retrieve the Tweets against selected hashtags. Tweet Archiver is an add-on used to save Tweets centred on popular hashtags, brand mentions, geo-tagged Tweets, keywords, and more. It polls Twitter every hour and extracts all matched Tweets into a Google Spreadsheet using basic searches, Boolean search, or sophisticated Twitter search operators to generate more complicated queries. The app is running in the background and auto-downloading the Tweets matching the query.

The search query employed in Tweet Archiver to retrieve the Tweets is

These hashtags: Covid19Vaccine, CoronaVirusVaccine.
Written in: English (The study included the Tweets that were posted in English language only)

The search query was created on 4 January 2021 and closed after 2 months 18 days (i.e. 22 March 2021). A total of 11,815 Tweets with accompanying metadata were retrieved for the selected period.

4.3. Phase III: data selection

A total of 11,815 Tweets were harvested based on the selected hashtags. The first step was to trace the exact location of the Tweet user in order to categorise the data according to the country. While categorising the data as per the location of the Twitter users, the majority of the users simply mentioned only the names of the cities and some specific area names. The names of the cities were individually searched on the Internet to locate their respective nations and were then labelled accordingly. Moreover, some Twitter users had mentioned their locations like Mars, above the sky, and Unknown and lacked information about the specific countries, and these Tweets were grouped under the Unknown category. After the whole data set was grouped according to their respective countries, the data pertaining to India were only selected for further analysis. There were 2700 Tweets related to India, and after thoroughly analysing these 2700 Tweets, 1269 duplicate retweets were found, which were excluded from the study, resulting in 1431 Tweets included for further analysis.

4.4. Phase IV: sentiment analysis of Tweets

To visualise the sentiments of the Tweets, the data set was migrated into the Orange Data mining Software. The corpus was pre-processed using the standard data pre-processing techniques, such as Transformation, Tokenization and Filtering using the Orange software. The 'Valence Aware Dictionary for sEntiment Reasoning' (VADER) was used to determine the sentiments of Tweets. Each word in the lexicon is automatically classified as positive, neutral or negative using this tool. The VADER is 'a rule-based model for general sentiment analysis. VADER is constructed using existing well-established sentiment lexicons using lexical features commonly used to express sentiment in social media and is currently considered as a gold standard in social media lexicons' [73]. VADER calculates the sentiments of each Tweet and returns a compound sentiment score ranging from − 1 to 1. Based on the classification thresholds established by the developers of the library,

> Positive sentiment was categorized using a sentiment score of greater than or equal to 0.05; sentiment scores between -0.05 and 0.05 were categorized as neutral, and the negative sentiment was represented using a sentiment score of less than or equal to 0.05.

A few previous studies [28,73,74] were found to be instrumental in this arena.

4.5. Phase V: word frequency

VOSviewer (a tool for creating and visualising bibliometric networks based on citation, bibliographic coupling, co-citation, or co-authorship connections) [75] version 1.6.17 (can be downloaded from: https://www.vosviewer.com) was used to illustrate the word frequency from the data set. VOSviewer has text mining capabilities for generating keyword co-occurrence networks from English-language textual data [76]. The Tweet data set was exported to VOSviewer, which extracted a total of 4689 terms, of which 127 terms were selected by using a threshold value of 10 minimum numbers of occurrences of a term. The relevance score was determined and 60% (i.e. 76) of the most relevant phrases were selected for analysis. The relevance score and occurrence of terms were recorded and a visualisation graph of the keywords was generated accordingly. Earlier studies have also used VOSviewer to analyse the Twitter data [77,78]. To learn how to utilise the tool step by step, novice researchers can consult [79].

5. Results

5.1. Geographical distribution of Tweets and sentiments of Tweets

Table 1, supplemented by Figures 1, 2, and 3, shows the geographical distribution of Tweets and their sentiment polarity. Figures 1, 2 and 3 represent the two shades, red and green. The red indicates the lowest number of Tweets while green shows the highest number of Tweets. The brighter the shade of green, the more productive the particular State is. During analysis, it is observed that the large number of Twitter users have not mentioned their exact state location; they have generally mentioned their location as India, Bharat, or Hindustan and the states which have posted less than 30 Tweets were kept in the 'others' category. The Tweets were posted across the country, with a large contribution from Maharashtra (223, 15.58%), followed by Delhi (220, 15.37%) and Tamil Nadu (73, 5.10%). However, the Uttar Pradesh, Karnataka and Telangana have posted more than 50 Tweets each while Gujarat, Jammu and Kashmir, Odisha and Punjab have posted in a range of 30–45 Tweets. The states that posted less than 30 Tweets about the vaccine constitute Rajasthan, West Bengal, Arunachal Pradesh, Bihar, Jharkhand, Madhya Pradesh, Assam, Haryana, Kerala, Chhattisgarh, Tripura, Himachal Pradesh, Nagaland, Bhutan, Meghalaya, Goa, Nepal, Pondicherry, Uttrakhand, Manipur and Andaman and Nicobar.

The detected Sentiments further supplement the findings by categorising Tweets into positive, negative, and neutral. It is apparent from Table 1 that the majority (639, 44.65%) of the Tweets reflect positive sentiments, followed by neutral sentiments (551, 38.50%), while the least proportion (241, 16.84%) of Tweets showcase a negative opinion about the vaccines. This signifies that most Twitter users have a positive outlook towards the Covid vaccines in India. The representative Tweets of positive, negative and neutral sentiments are provided in Table 2.

6. Word frequency of frequently appearing terms in Tweets based on their relevance score

Table 3 lists the 76 most frequently occurring terms used by Indians to express their opinions about the Covid-19 vaccination. The words 'Delhi heart', 'Lung institute', 'Gift', 'Unite2fightcorona', and 'Covid19 Vaccine' are the leading

Table 1. Top 10 tweeting states and their sentiments.

Rank	State	Tweet count	% age (n = 1431)	Sentiments					
				Positive	%age	Negative	%age	Neutral	%age
1	Maharashtra	223	15.58	101	45.29	38	17.04	84	37.66
2	Delhi	220	15.37	87	39.54	40	18.18	93	42.27
3	Tamil Nadu	73	5.10	29	39.72	6	8.21	38	52.05
4	Uttar Pradesh	64	4.47	31	48.43	9	14.06	24	37.5
5	Karnataka	54	3.77	30	55.55	5	9.25	19	35.18
6	Telangana	50	3.49	20	40	7	14	23	46
7	Gujarat	45	3.14	26	57.77	5	11.11	14	31.11
8	Jammu & Kashmir	37	2.58	14	37.83	9	24.32	14	37.83
9	Odisha	36	2.51	9	25	4	11.11	23	63.88
10	Punjab	32	2.23	18	56.25	2	6.25	12	37.5
	Others[a]	597	41.71	274	45.89	116	19.43	207	34.67
		1431	100	639	44.65	241	16.84	551	38.50

[a]The Tweets related to those users who have not mentioned the exact state location; they have generally mentioned their location as India, Bharat and Hindustan so authors were not able to categorise them as per state or Tweets related to those states which have posted less than 30 Tweets.

Table 2. Representative Tweets of positive negative and neutral sentiments.

Positive Tweets	Negative Tweets	Neutral Tweets
1. 'some of the college staff members got first dose of covid-19 vaccine covishield i appeal to all to take the vaccine together let us make india covid-19 free and defeat thisvirus. surat covid19vaccine indiafightscorona largestvaccinedrive covid19'	1. 'canada suspended the use of the oxford astrazeneca coronavirusvaccine for people under age 55 following concerns it might be linked to rare blood clots'	1. '... The Rise of NaMo and New India' charts India?s 7 year history of Modi rule. It examines critically the achievements of his government. It would also give readers new insights into political conundrum. NamoAndNewIndia #Modiji'
2. 'today i received my first dose of the covid19vaccine i encourage everyone to book an appointment as soon as it's your turn'	2. 'the virus is still rampaging around most of the planet and uneven vaccine distribution poses a major public risk as variants emerge coronavirus covid19 coronavirusvaccine'	2. 'the virus is still rampaging around most of the planet and uneven vaccine distribution poses a major public risk as variants emerge coronavirus covid19 coronavirusvaccine'
3. 'drive away the fear of coronavirus with the covid19 vaccination and enjoy the colourful spirit of holi be responsible, stay safe vaccinewaliholi holi holi2021 covid19vaccine watergun colours kojakselinge coronavirus hmdfightscorona topicalpost'	3. '... refuses to take bjp's covid-19 vaccine 178 detainees in up prison get inoculated covid19vaccine azamkhan uttarpradesh trhnews therealityhunt therealityhuntnews therealityhuntlive therealityhuntlivenews'	3. '...what concentration of antibodies may be needed to provide such protection covid19 covid19 covid19 covid covid19 corona coronaviruspandemic covidvaccine covid19vaccine covidvaccination covid19vaccination 2/2'

words with the highest relevance score ranging between 2.91 and 1.82. The terms with higher relevance scores represent specific topics covered by the test data, while the terms with lower relevance scores are general and not representative of any specific topic [71,80]. However, based on the occurrence of terms, the terms 'Year', 'April', 'Covidshield', 'Week', and 'age' have occurred more than 50 times, but their relevance score is less, which ranges between '0.43 and 0.65'. A graphic depiction of terms appearing in Tweets is displayed in Figure 4. The bigger the keyword, the greater its occurrence counts in the Tweets is.
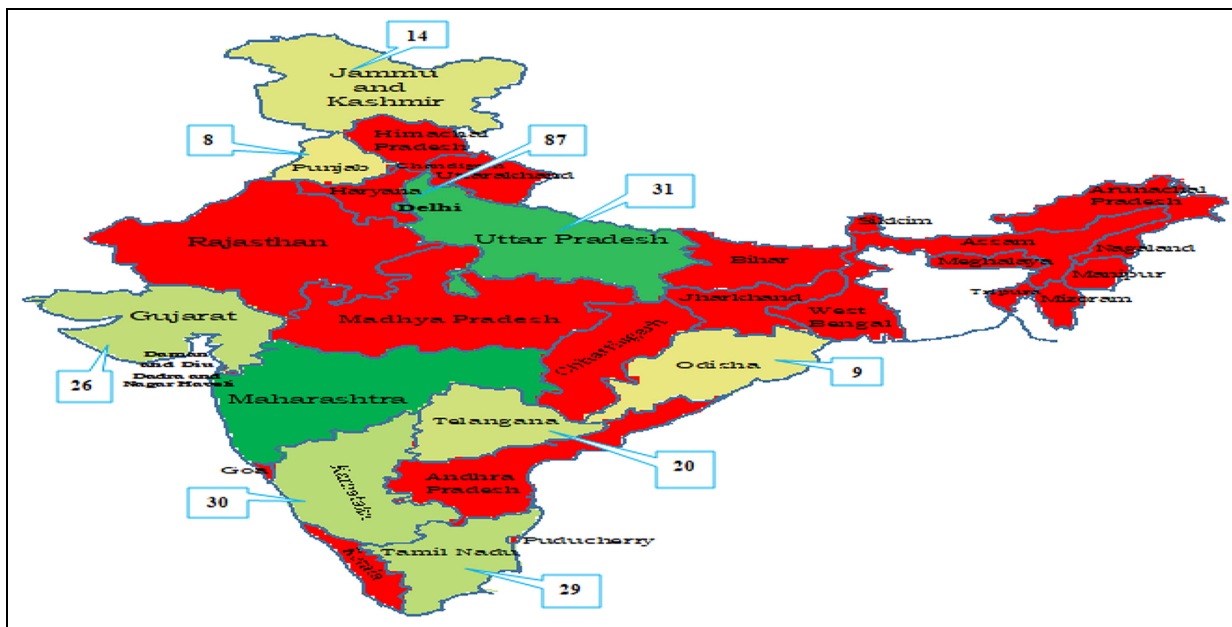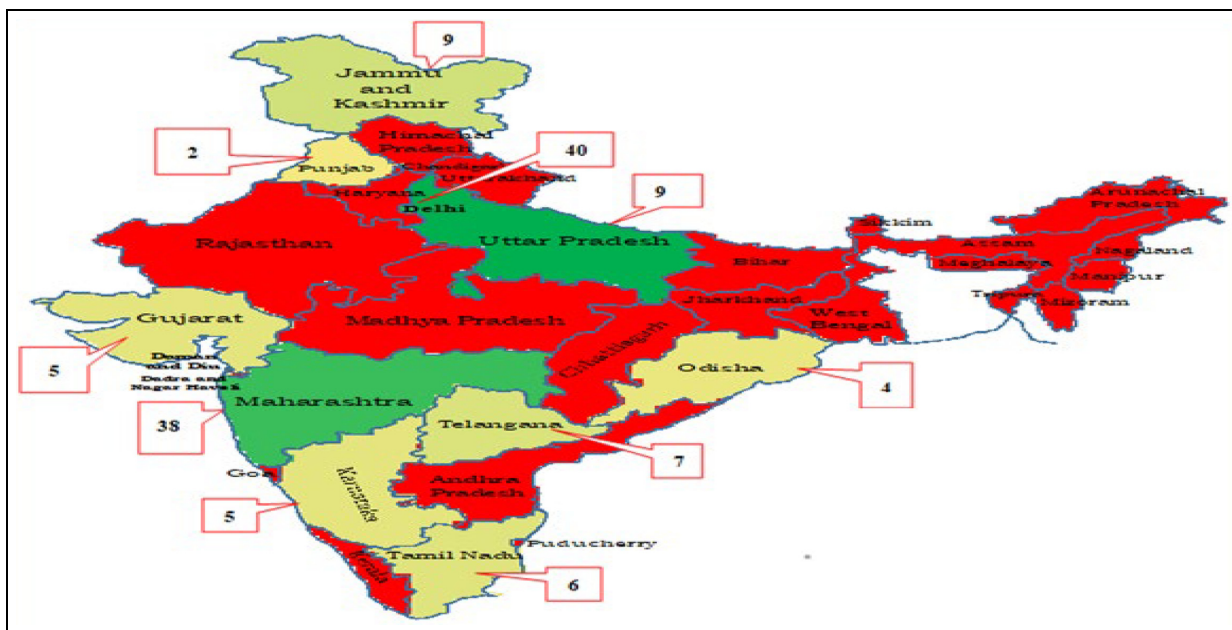
Figure 1. Positive sentiment Tweets.
The figure depicts the vivid picture of states and their sentiments. The template for geographical heat map was downloaded from https://indzara.com and loaded with own data set.

Figure 2. Negative sentiment Tweets.
The figure depicts the vivid picture of states and their sentiments. The template for geographical heat map was downloaded from https://indzara.com and loaded with own data set.
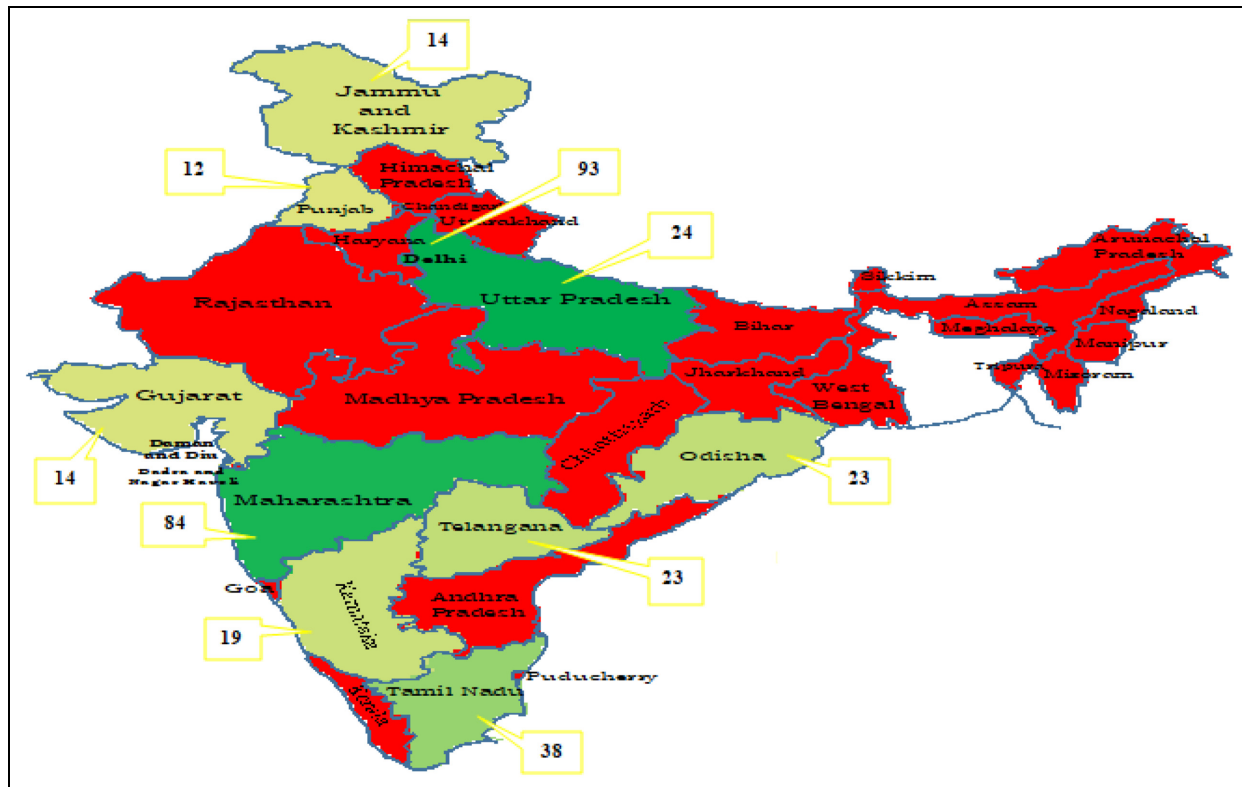
Figure 3. Neutral sentiment Tweets.
The figure depicts the vivid picture of states and their sentiments. The template for geographical heat map was downloaded from https://indzara.com and loaded with own data set.

7. Cluster analysis of frequently appearing terms in Tweets

Based on the VOSviewer analysis, the terms were grouped into eight clusters, of which the first three clusters consisted of more than 10 items each as depicted in Table 3 and Figure 4. Cluster 1 (red cluster), located in the lower-left area in the visualisation graph, consists of terms related to the covidshield vaccine and the coronavirus update. The top five terms based on their total link strength (TLS) are Covidshield, Coronavirus update, life, work and Covid19vaccine date in this cluster. Cluster 2 (green cluster), located in the upper right area, covers terms related to coronavirus, the Indian Covid seva platform and the appointment for the Covid vaccination dose. The top five prominent terms in this cluster based on their TLS are corona coronavirus, Covid indiaseva, Covidvaccination, appointment and second dose. Cluster 3 (sky blue cluster), positioned in the left middle corner, comprised the terms related to Covid Tests, new cases, and lockdown. Based on the TLS of words, the top five terms are Year, Covid-19 India, date, lockdown, total tests and new cases. Cluster 4 (yellow cluster), found in the middle area of the visualisation graph, is related to the terms that report the Covid deaths in different states of India. The top five terms in this cluster based on TLS are Maharashtra, Punjab, report, death and vaccine dose. Cluster 5 (violet cluster), located in the bottom corner of the visualisation graph, is related to the terms related to the AstraZeneca vaccine trials and their efficiency. The top words in this cluster are AstraZeneca, efficacy, trial and study. Cluster 6 (blue cluster), located in the lower right area in the visualisation, is related to the terms discussing the gap/interval between the vaccine doses for protection. The top five terms based on TLS are Week, interval, centre, gap and protection. Cluster 7 (orange cluster), placed in the top left corner, is associated with terms that address citizens' age and comorbidity. This cluster's leading terms based on TLS are April, age, citizen and comorbidity. Cluster 8 (brown cluster) is placed in the right-centre corner region and contains terms related to the inoculation of Covid vaccination doses at the Delhi Heart and Lung Institute. The group comprises the terms second dose, lung institute, Delhi heart and Delhi.

Table 3. Displaying the clusterwise top frequently appearing keywords in Tweets with their links and total link strength.

Cluster	No. of terms	Links of whole cluster	Total link strength of whole cluster	Terms
1	21	250	545	Astrazenecacovid19vaccine, Bhutan, China, Corona, Coronavirusupdate, Covid19 Vaccine, Covid19vaccineDose, Covishield, Doctor, Gift, Holi, Home, Indian, Life, Mask, Pune, Read, Russia, Took, Turn, Work
2	12	218	659	1st Dose, 2nd Dose, Appointment, Book, Co morbidity, Contact, Corona Coronavirus, Covid Indiaseva,Covid19vaccination Centre, Rt Covid Indiaseva, Vaccination Centre
3	10	167	773	Covid19india, Covidvaccine Covid19vaccine, Crore, Date, Israel, Lockdown, New Cases, Total Cases, Vaccinationdrive, Year
4	9	145	332	Crcovid19vaccine, Death, Hour, Karnataka, Maharashtra, Punjab, Report, Unite2fightcorona, Vaccine Dose
5	8	90	212	Astrazeneca, Company, Efficacy, Lakh Dose, Rt Indiatoday, Serum Institute, Study, Trial,
6	7	129	499	Center, Covishield Vaccine, Gap, Interval, Odisha, Protection, Week
7	5	89	383	Age, April, Citizen, Co Morbidity, Everybody
8	4	47	187	Delhi, Delhi Heart, Lung Institute, Second Dose

8. Discussion and findings

People are increasingly utilising the Internet and social media platforms for seeking health-related information and keeping track of public sentiments on controversial issues like vaccination [81]. This study focuses on the sentiments and frequent keyword expressions of Indian Tweets related to the 'Covid-19 vaccines' discourse on Twitter. The Twitter data were utilised because user ideas linked with certain events and features describing events according to user perception may be discovered using Twitter data [26]. The 'Twitter platform used in the present study can be a valuable tool for public health promotion to reinforce vaccine acceptance and decrease vaccine hesitancy and opposition'[21]. It is observed from the study that the Tweets posted across the country with the highest contribution were found to be from Maharashtra (223, 15.58%), followed by Delhi (220, 15.37%) and Tamil Nadu (73, 5.10%), respectively. The highest contribution from Maharashtra can be attributed to a larger population of the state with many Twitter users, as Maharashtra is the second largest populated state in India. The sentiment analysis of the Tweets highlights the considerable difference in the prevalence of positive, negative and neutral sentiments. The result shows that most of the Tweets have positive opinions about Covid-19 vaccines while the least showcase negative sentiments. The findings of the current study align with the earlier studies [19,28,46,55,57,61,66,70] signifying that positive sentiments about Covid-19 vaccines are dominant on Twitter. The results indicate that most Twitter users are hopeful and positive about Covid-19 vaccines in India. However, negative sentiments have also remained in the least proportion of people, increasing concerns about getting the Covid-19 vaccination. The negativity about the Covid-19 vaccine can be eliminated from people's minds by spreading awareness about the benefits of vaccination [28] and the influence of disinformation about vaccination can be mitigated by identifying people's general opinions [21]. Based on the relevance score, it is observed that the word 'Delhi heart', 'Lung institute', 'Gift', 'Unite2fightcorona' and 'Covid19 Vaccine' are the leading words people have used to share their thoughts about Covid-19 vaccines in India. In early 2021, the devastating second wave of the Covid-19 epidemic struck havoc in Delhi, claiming many lives and taxing hospitals and physicians beyond their
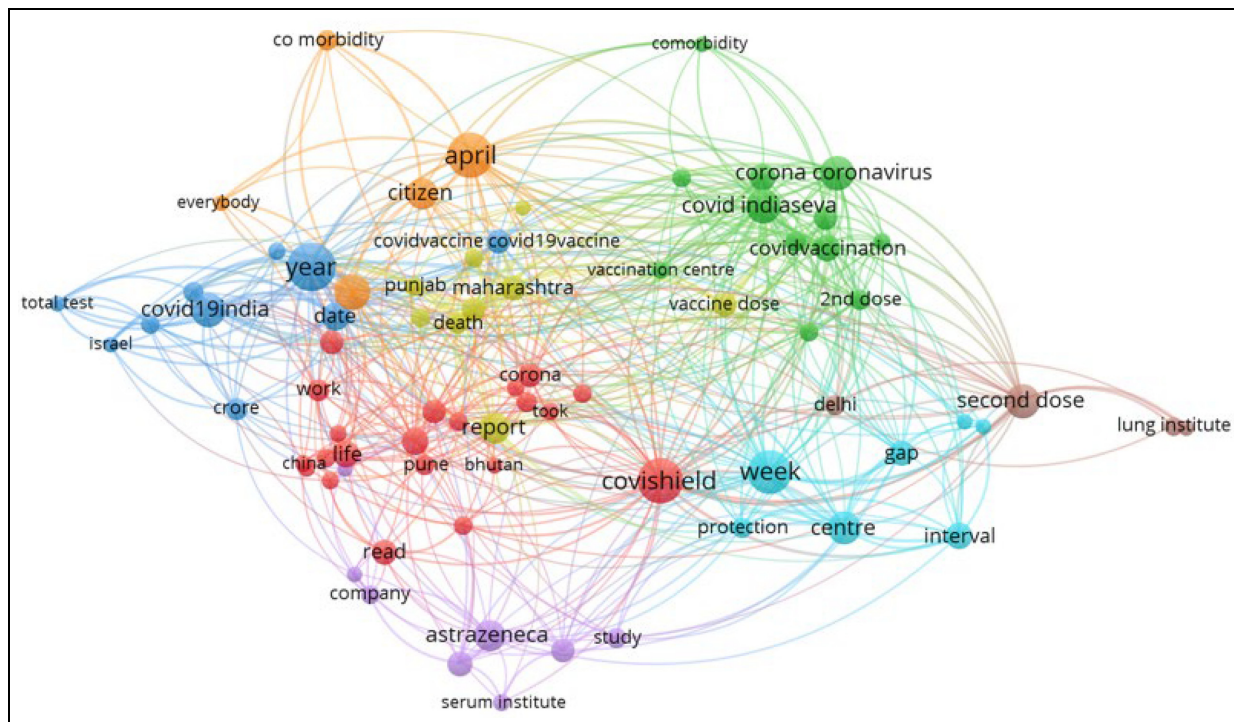
Figure 4. Visualisation graph of most frequently used terms in Tweets.

capacity [82]. Many people died in India due to the spreading of infection in the lungs. So it is a natural phenomenon that the affected people will prefer to go for treatment at specialty institutes/hospitals. So, some people have expressed that they had been inoculated with vaccination at Delhi Heart & Lung Institute, which might be the possible reason why the words 'Delhi heart' and 'Lung institute' often appear in Tweets, for example, '... Watch Now ! Union Health Minister Dr Harsh Vardhan gets inoculated with 2nd dose of #COVID19Vaccine at Delhi Heart & Lung Institute @PMOIndia @MoHFW_INDIA #LargestVaccineDrive'. The popularity of the term 'gift' can be ascribed to the fact that India, as one of the world's top vaccine makers, provided the Covid-19 vaccine as a gift to many nations throughout the globe, and the same is reflected by Twitter users in their Tweets, for example, '... Top UN officials, including the UN peacekeeping chief, have expressed gratitude to India for its gift of 2,00,000 #COVID19Vaccine doses for peace-keepers, saying the donation will enable the Blue Helmets to continue their life-saving work in a safe manner'. However, the phrase 'unite2fightcorona' gained popularity as a result of most Twitter users using it as a hashtag to share their thoughts on the vaccination and to rally others to fight the epidemic, for example, 'Let us join the fight to make India & world #COVID19 free Together we can & together we will! #Unite2FightCorona #IndiaFightsCorona #VaccinationDrive'. The prominence of the term 'covid 19 vaccination' can be attributed to the fact that the whole discourse is focused on it and data were also obtained using the same hashtag.

9. Conclusion

This study highlights the common people's sentiments towards the Covid-19 vaccines in India. The study shows that Twitter has evolved into a vital source of information and can be utilised in real-time situations to manage various crises by properly analysing the data. By identifying the sentiments, public reactions about the Covid-19 vaccine can be determined, which will aid in developing programmes for providing authoritative health information and improving communication to foster understanding and confidence [83]. However, it is disconcerting that the negative sentiment towards vaccines rests with some people in India. So, public healthcare agencies need to focus on social media data to promote positive messaging and decrease negative views [84]. The Twitter data utilised in this study could be an effective strategy for public health awareness. The study suggests that, on days when a serious incident or crisis happens, government administrators and policymakers should pay attention to and monitor the dialogues on Twitter to establish appropriate

strategies to support public communication to minimise the damage/outbreak, since effective crisis management and communication can help reduce damages, uncertainty and contribute to a quick recovery [85].

10. Practical implications

The study outcomes will be used to understand the opinions of Indian people about the Covid-19 vaccine, which may help public health authorities understand the polarity (positive, negative, and neutral) Tweets about Covid-19 vaccines on Twitter. The research might be helpful for health policymakers and administrators to enhance public awareness about health concerns and misinformation about Covid-19 vaccine. The study can also assist the healthcare sector in accentuating positive messages and reducing negative ones to enhance vaccination uptake. The study provides useful insights into how online studies related to social media data can be structured for information extraction and analysis in order to gain fascinating insights into the Covid-19 vaccinations and other scenarios.

11. Limitation and future research

The study cannot be generalised to the larger population because everybody is not using Twitter. Other social media platforms can also be used for a comprehensive understanding of the viewpoints expressed by people on various social media platforms. The limited number of Tweets with a specific hashtag were considered for the study and it is obvious that the Tweets that were posted under the other hashtags were not entertained in the study. Because of the manual examination of the Tweets, the research only included a small number of Tweets for analysis. It would be fascinating to examine the large-scale data in light of other popular hashtags associated with the Covid-19 vaccines, which may provide different valuable findings in future research. The study only looks at Tweets from a single nation. Future research should include Tweets from other nations for comparative sentiment analysis. Furthermore, the study looks at Tweets written in the English language only. It could be interesting to incorporate Tweets in other languages in future study to emphasise the diverse opinions of the people. Tweet Archiver was used to store Tweets focusing on prominent hashtags; the add-on has some limitations. Researchers can utilise the Twitter application programme interface (API) to retrieve data in the future.

ORCID iD

Aasif Ahmad Mir https://orcid.org/0000-0002-0492-728X

References

[1] Yan Y, Shin WI, Pang YX et al. The first 75 days of novel coronavirus (SARS-CoV-2) outbreak: recent advances, prevention, and treatment. *Int J Environ Res Public Health* 2020; 17(7): 2323.
[2] Paules CI, Marston HD and Fauci AS. Coronavirus infections – more than just the common cold. *JAMA* 2020; 323(8): 707–708.
[3] Richman D, Whitley J, Hayden G et al. *Clinical virology*. 4th ed. Washington, DC: ASM Press, 2016.
[4] Weiss SR and Leibowitz JL. Coronavirus pathogenesis. *Adv Virus Res* 2011; 81: 85–164.
[5] Masters PS and Perlman S. Coronaviridae. In: Knipe DM and Howley PM (eds) *Fields virology*. 6th ed. London: Lippincott Williams & Wilkins, 2013, pp. 825–858.
[6] Su S, Wong G, Shi W et al. Epidemiology, genetic recombination, and pathogenesis of coronavirus. *Trend Microbiol* 2016; 24(6): 490–502.
[7] Feng Z, Yu Q, Yao S et al. Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics. *Nat Commun* 2020; 11(1): 4968.
[8] Gandhi RT, Lynch JB and Del R. Mild or moderate covid-19. *N Eng J Med* 2020; 383(18): 1757–1766.
[9] World Health Organization. Health topics/coronavirus, 2020, https://www.who.int/health-topics/coronavirus (accessed 28 March 2021).
[10] Zaim S, Chong JH, Sankaranarayanan V et al. COVID-19 and multiorgan response. *Curr Prob Cardiol* 2020; 45(8): 100618.

[11] Huang C, Wang Y, Li X et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395(10223): 497–506.

[12] Zhu N, Zhang D, Wang W et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Eng J Med.* 2020; 382: 727–733.

[13] Cheng A and Williamson D. An outbreak of COVID-19 caused by a new coronavirus: what we know so far. *Med J Australia* 2020; 212(9): 393–394.

[14] *Report of clustering pneumonia of unknown etiology in Wuhan City.* Wuhan, China: Wuhan Municipal Health Commission, 2019.

[15] Coronavirus disease, https://www.who.int/india/emergencies/coronavirus-disease-(covid-19) (2020, accessed 30 July 2021).

[16] Lai CC, Shih TP, Ko WC et al. Acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int J Antimicrob Agents* 2020; 55(3): 105924.

[17] Liu YC, Kuo RL and Shih SR. COVID-19: the first documented coronavirus pandemic in history. *Biomed J* 2020; 43(4): 328–333.

[18] WHO. Director-General's opening remarks at the media briefing on COVID-19, https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19––1-april-2020 (2020, accessed 1 April 2020).

[19] Dua S. Sentiment analysis of COVID-19 vaccine Tweets, https://towardsdatascience.com/sentiment-analysis-of-covid-19-vaccine-Tweets-dc6f41a5e1af (2021, accessed 21 January 2022).

[20] Jayasurya GG, Kumar S, Singh BK et al. Analysis of public sentiment on COVID-19 vaccination using Twitter. *IEEE Trans Comput Soc Syst* 2021; 9: 1101–1111.

[21] Yousefinaghani S, Dara R, Mubareka S et al. An analysis of COVID-19 vaccine sentiments and opinions on twitter. *Int J Infect Dis* 2021; 108: 256–262.

[22] Love B, Himelboim I, Holton A et al. Twitter as a source of vaccination information: content drivers and what they are saying. *Am J Infect Control* 2013; 41(6): 568–570.

[23] Coronavirus disease 2019. Center for Disease Control and Prevention, https:/www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html (2020, accessed 21 January 2022).

[24] Carley KM, Malik M, Landwehr PM et al. Crowd sourcing disaster management: the complex nature of Twitter usage in Padang Indonesia. *Saf Sci* 2016; 90: 48–61.

[25] McNely B. Backchannel persistence and collaborative meaning making. In: *SIGDOC'09: proceedings of the 27th ACM international conference on design of communication*, Bloomington, IN, 5–7 October 2009, pp. 297–304. New York: ACM.

[26] Baralis E, Cerquitelli T, Chiusano S et al. Analysis of twitter data using a multiple-level clustering strategy. In: *International conference on model and data engineering*, September 2013, pp. 13–24. Berlin; Heidelberg: Springer, https://core.ac.uk/download/pdf/18304796.pdf

[27] Fausto S and Aventurier P. Scientific Literature on Twitter as a subject research: findings based on bibliometric analysis, 2016, https://hal.archives-ouvertes.fr/hal-01297804

[28] Mir AA, Rathinam S and Gul S. Public perception of COVID-19 vaccines from the digital footprints left on Twitter: analyzing positive, neutral and negative sentiments of Twitterati. *Libr Hi Tech* 2021; 42(2): 340–356.

[29] Jordan SE, Hovet SE, Fung ICH et al. Using Twitter for public health surveillance from monitoring and prediction to public response. *Data* 2018; 4(1): 6.

[30] Sinnenberg L, Buttenheim AM, Padrez K et al. Twitter as a tool for health research: a systematic review. *Am J Public Health* 2017; 107(1): e1–e8.

[31] Shamrat MFMJ, Chakraborty S, Imran MM et al. Sentiment analysis on twitter Tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. *Indones J Electr Eng Comput Sci* 2021; 23(1): 463–470.

[32] Tomeny TS, Vargo CJ and El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. *Soc Sci Med* 2017; 191: 168–175.

[33] Velasco E, Agheneza T, Denecke K et al. Social media and internet-based data in global systems for public health surveillance: a systematic review. *Milbank Quart* 2014; 92(1): 7–33.

[34] Yousefinaghani S, Dara R, Poljak Z et al. The assessment of Twitter's potential for outbreak detection: avian influenza case study. *Sci Rep* 2019; 9(1): 1–17.

[35] Ardianto R, Rivanie T, Alkhalifi Y et al. Sentiment analysis on E-sports for education curriculum using Naive Bayes and support vector machine. *Jurnal Ilmu Komputer Dan Informasi* 2020; 13(2): 109–122.

[36] Yan SA and Mawhorter P. Twitter sentiment analysis: fan engagement in esports matches, 2020, https://www.elearning-conf.org/wp-content/uploads/2020/07/03_202008C035_S049.pdf

[37] Zhao S, Zhong L, Wickramasuriya J et al. Analyzing twitter for social TV: sentiment extraction for sports. *Proc Int Work Future Telev* 2011; 2: 11–18.

[38] Yu Y and Wang X. World Cup 2014 in the Twitter World: a big data analysis of sentiments in US sports fans' Tweets. *Comput Hum Behav* 2015; 48: 392–400.

[39] Schumaker RP, Jarmoszko AT and Labedz CS. Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decis Support Syst* 2016; 88: 76–84.

[40] Wunderlich F and Memmert D. Innovative approaches in sports science – Lexicon-based sentiment analysis as a tool to analyze sports-related Twitter communication. *Appl Sci* 2020; 10(2): 431.

[41] Hegde SU, Zaiba AS and Nagaraju Y. Hybrid CNN-LSTM model with GloVe word vector for sentiment analysis on football specific Tweets. In: *ICAECT*, Bhilai, India, 19–20 February 2021, pp. 1–8. New York: IEEE.

[42] Gupta VS and Kohli S. Twitter sentiment analysis in healthcare using Hadoop and R. In: *2016 3rd international conference on computing for sustainable global development (Indiacom)*, New Delhi, India,16–18 March 2016, pp. 3766–3772. New York: IEEE.

[43] Gohil S, Vuik S and Darzi A. Sentiment analysis of health care Tweets: review of the methods used. *JMIR Pub Health Surveill* 2018; 4(2): e5789.

[44] Clark EM, James T, Jones CA et al. A sentiment analysis of breast cancer treatment experiences and healthcare perceptions across twitter. *Arxiv Preprint*, 2018, https://arxiv.org/abs/1805.09959

[45] Beigi G, Hu X, Maciejewski R et al. An overview of sentiment analysis in social media and its applications in disaster relief. *Sent Anal Ontol Eng* 2016; 2016: 313–340.

[46] Vera VDG, Suárez LMM and Lopera ICP. Sentiment analysis on post conflict in Colombia: a text mining approach. *Asian J Appl Sci* 2018; 6(2): 885.

[47] Pengnate SF, Lehmberg DG and Tangpong C. Top management's communication in economic crisis and the firm's subsequent performance: sentiment analysis approach. *Corporate Commun* 2020; 25(2): 187–205.

[48] Xiong J, Hswen Y and Naslund JA. Digital surveillance for monitoring environmental health threats: a case study capturing public opinion from Twitter about the 2019 Chennai water crisis. *Int J Environ Res Public Health* 2020; 17(14): 5077.

[49] Bashir S, Bano S, Shueb S et al. Twitter chirps for Syrian people: sentiment analysis of Tweets related to Syria Chemical attack. *Int J Disaster Risk Reduct* 2021; 62: 102397.

[50] Tian X, He W and Wang FK. Applying sentiment analytics to examine social media crises: a case study of United Airline's crisis in 2017. *Data Technol Appl* 2021; 56(1): 1–23.

[51] Marcec R and Likic R. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgrad Med J* 2021; 98: 140685.

[52] Lyu JC, Le HE and Luli GK. COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis. *J Medical Internet Res* 2021; 23(6): e24435.

[53] Sattar NS and Arifuzzaman S. COVID-19 Vaccination awareness and aftermath: public sentiment analysis on Twitter data and vaccinated population prediction in the USA. *Appl Sci* 2021; 11(13): 6128.

[54] Ansari M and Khan N. Worldwide COVID-19 vaccines sentiment analysis through Twitter content. *Electron J General Med* 2021; 18(6): 11316.

[55] Liu S and Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: a sentiment analysis. *Vaccine* 2021; 39: 5499–5505.

[56] Hu T, Wang S, Luo W et al. Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: spatiotemporal perspective. *J Med Internet Res* 2021; 23(9): e30854.

[57] Khakharia A, Shah V and Gupta P. Sentiment analysis of COVID-19 vaccine Tweets using machine learning, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3869531

[58] Nuzhath T, Tasnim S, Sanjwal RK et al. COVID-19 vaccination hesitancy, misinformation and conspiracy theories on social media: a content analysis of Twitter data, 2020, https://osf.io/preprints/socarxiv/vc9jb/

[59] Griffith J, Marani H and Monkman H. COVID-19 vaccine hesitancy in Canada: content analysis of Tweets using the theoretical domains framework. *J Med Int Res* 2021; 23(4): e26874.

[60] Bonnevie E, Gallegos JA, Goldbarg J et al. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. *J Commun Healthc* 2021; 14(1): 12–19.

[61] Thelwall M, Kousha K and Thelwall S. Covid-19 vaccine hesitancy on English-language Twitter. *Prof De La Inf (EPI)* 2021; 30(2): 3145.

[62] Jemielniak D and Krempovych Y. An analysis of AstraZeneca COVID-19 vaccine misinformation and fear mongering on Twitter. *Public Health* 2021; 200: 4–6.

[63] Hayawi K, Shahriar S, Serhani MA et al. ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public Health* 2022; 203: 23–30.

[64] Sharma K, Zhang Y and Liu Y. COVID-19 vaccines: characterizing misinformation campaigns and vaccine hesitancy on twitter. *Arxiv Preprint* 2021; 2021: 2106.08423.

[65] Trisnawati S, Priantoro AU, Fauzi C et al. Twitter sentiment analysis on the use of Sinovac Vaccine in Indonesia. In: *ICITB*, 2021, December, pp. 77–80, https://jurnal.darmajaya.ac.id/index.php/icitb/article/view/3047

[66] Aryal RR and Bhattarai A. Sentiment analysis on Covid-19 vaccination Tweets using Naïve Bayes and LSTM. *Adv Eng Technol* 2021; 1(1): 57–70.

[67] Dubey AD. Public sentiment analysis of COVID-19 vaccination drive in India, https://ssrn.com/abstract=3772401 (accessed 24 January 2021).

[68] Alselwia G and Kaynak S. *Sentiment analysis on Covid-19 vaccines Tweets*. International Marmara Sciences Congress (Spring), 2021, 2021, https://www.researchgate.net/profile/Ghadir-Alselwi/publication/354223609_SENTIMENT_ANALY

SIS_ON_COVID-19_VACCINES_TWEETS/links/612d1cba0360302a006c6085/SENTIMENT-ANALYSIS-ON-COVID-19-VACCINES-TWEETS.pdf

[69]   Hussain A, Tahir A, Hussain Z et al. Artificial intelligence–enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. *J Med Internet Res* 2021; 23(4): e26627.

[70]   Mahanti A, Chakraborty SD, Goel S et al. Are you positive about Covid vaccination? A comparative study of Twitter Sentiment, https://www.researchgate.net/profile/Shreyan-Chakraborty/publication/356612034_Are_you_Positive_about_Covid_Vaccination_A_Comparative_Study_of_Twitter_Sentiment/links/61a4d21d8c253c45f695e6a3/Are-you-Positive-about-Covid-Vaccination-A-Comparative-Study-of-Twitter-Sentiment.pdf (2021, accessed 30 January 2022).

[71]   Kwok SWH, Vadde SK and Wang G. Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: machine learning analysis. *J Med Internet Res* 2021; 23(5): e26953.

[72]   Gulati S. Decoding the global trend of 'vaccine tourism' through public sentiments and emotions: does it get a nod on Twitter. *Glob Knowl Mem Commun*. Epub ahead of print 9 September 2021. DOI: 10.1108/GKMC-06-2021-0106.

[73]   Bonta V and Janardhan NKN. A comprehensive study on lexicon based approaches for sentiment analysis. *Asian J Comput Sci Technol* 2019; 8(S2): 1–6.

[74]   Hung M, Lauren E, Hon ES et al. Social network analysis of COVID-19 sentiments: application of artificial intelligence. *J Medical Internet Res* 2020; 22(8): e22590.

[75]   VOSviewer. VOSviewer – visualizing scientific landscapes, https://www.vosviewer.com/ (2022, accessed 3 March 2022).

[76]   Waltman L, van Eck NJ and Noyons ECM. A unified approach to mapping and clustering of bibliometric networks. *J Informetr* 2010; 4(4): 629–635.

[77]   Yoon YS, Oh HW and Park K. Mining twitter to identify customers' requirements and shoe market segmentation. In: *International conference on information and communication technology convergence (ICTC)*, Jeju, South Korea, 17–19 October 2018, pp. 1194–1199. New York: IEEE.

[78]   Noor S, Guo Y, Shah SHH et al. Analysis of public reactions to the novel Coronavirus (COVID-19) outbreak on Twitter. *Kybernetes* 2020; 50(5): 1633–1653.

[79]   Eck NJV and Waltman L. Visualizing bibliometric networks. In: Ding T, Rousseau R and Waltman D (eds) *Measuring scholarly impact: methods and practice*. Cham: Springer, 2014, pp. 285–320.

[80]   Van Eck NJ and Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* 2017; 111(2): 1053–1070.

[81]   Brunson EK. The impact of social networks on parents' vaccination decisions. *Pediatrics* 2013; 131(5): e1397–e1404.

[82]   The Economic Times. News 2nd wave of Covid wreaked havoc in Delhi in 2021; Omicron shadow in 2022, https://economictimes.indiatimes.com//news/india/2nd-wave-of-covid-wreaked-havoc-in-delhi-in-2021-omicron-shadow-in-2022/articleshow/88633932.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst (2022, accessed 2 May 2022).

[83]   Monselise M, Chang CH, Ferreira G et al. Topics and sentiments of public concerns regarding COVID-19 vaccines: social media trend analysis. *J Medical Internet Res* 2021; 23(10): e30765.

[84]   Nezhad ZB and Deihimi MA. Twitter sentiment analysis from Iran about COVID 19 vaccine. *Diabet Metabol Syndr* 2022; 16(1): 102367.

[85]   Rodriguez TD and Dominguez LY. Social media in crisis communication: Germanwings Flight 4U9525. In: Freire F, Rúas Araújo X, Martínez Fernández V et al. (eds) *Media and metamedia management. Advances in intelligent systems and computing*, vol. 503. Cham: Springer, 2017, pp. 243–249.