Communications in Mathematics and Applications
Vol. 13, No. 1, pp. 379–386, 2022
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
DOI: 10.26713/cma.v13i1.1977

http://www.rgnpublications.com

Review Article

Analyzing Public Concerns Over COVID-19 Variants Using Social Media

Musleh Alsulami

Department of Information Systems, Umm Al-Qura University, Saudi Arabia
mhsulami@uqu.edu.sa

Received: January 12, 2022 **Accepted:** March 25, 2022

Abstract. SARS-CoV-2, or more popularly known COVID-19 has claimed more than 5.5 million lives since it has been declared as a global pandemic. Similar to other viruses, COVID-19 is also undergoing several mutations and has many variants like Alpha, Beta, Gamma, Delta, Omicron and others. With so many variants, social media users are confused and posting their frustrations and angers with Tweets or Posts in public social media platforms. These publicly accessible social media posts provide a wealth of information for a social scientist or political leader or a strategic decision maker. This study demonstrates a feasible approach to extract meaningful critical information from social media posts. By programmatically accessing Twitter database from 11th January 2022 till 20th January 2022, we retrieved almost 9 K Tweet messages on 6 different keywords like "COVID Variants", "Omicron", "Alpha Variant", "Beta Variant", "Gamma Variant" and "Delta Variant". Results were compared against metrics like users, posts, engagement, and influence. Omicron was found to be the most popular topic compared to other variants with an influence score of 70.2 million and 2.1 K posts during the monitored period. The most popular sources for influences on COVID-19 Variant related posts were found to be @reuters with 24.2M, @forbes with 17.4M, @timesofindia with 14.2M and @inquirerdotnet with 3.4 followers. This study also found out that the most popular Tweet languages were English followed by French and Dutch. Lastly, this study ranked user mentions, word frequency (with word cloud) and hashtags for COVID-19 Variant related twitter posts during the monitored timeframe.

Keywords. Social media analysis, Tweet analysis, Covid tweet, Covid hashtag, Number of posts, Virus expression, Viruses and cancer, Viral diseases

Mathematics Subject Classification (2020). 68R01, 57Z25

1. Introduction

COVID-19 or SARS-CoV-2 was first detected in December 2019 and on 11 March 2020 it was declared as a global pandemic by World Health Organization (WHO) [13]. At the time of conducting this study, COVID-19 has claimed more than 5.5 million lives making this pandemic as one of the worst pandemics the humankind has ever witnessed [13]. Since the beginning of inception COVID-19 has mutated into several variants, commonly named as Alpha, Beta, Gamma, Delta and more recently Omicron [14]. These mutations or changes include characteristics like how easily it spreads, the associated disease severity, or the performance of vaccines, therapeutic medicines, diagnostic tools, or other public health and social measures [14]. Other than Alpha, Beta, Gamma, Delta and Omicron, there are several other variants being reported on a regular basis [14]. Alpha variant is known as B1.1.1.7. It was first detected in United Kingdom in November 2020. It has symptoms like Chills, Loss of appetite, Headaches, Muscle ache etc. Beta Variant is also known as B.1.351. It was first detected in Nelson Mandela Bay, South Africa and reported in July 2020. Gamma Variant is also known as P.1. It was first detected in Tokyo, Japan. It was first reported on 6 January 2021. Delta Variant is also known as B.1.617.2. It was first detected on India. It was first reported on 2020. Omicron Variant is also known as B1.1.529. It was first detected in South Africa on 24 November 2021. Omicron has symptoms like Fatigue, body aches, Cough, Congested or running nose, Headache, Night Swear etc.

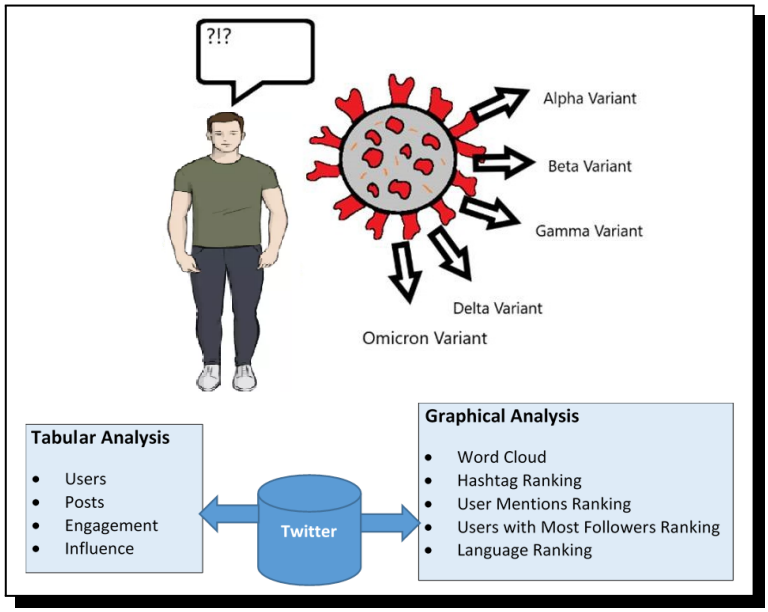


Figure 1

As seen from Figure 1, social media users are constantly posting their confusion, concerns, frustration, and anger over different variants of SARS-CoV-2. According to [1], out of the 7.89 billion world population, 5.29 billion users own mobile phones, and 4.88 billion users have internet connection in 2021. Most importantly, 4.55 billion users are classified as active social media users [1]. Since more than 57.6% of the world populations are now active social media user, tapping the social media data could provide a wealth of information for social scientists and researchers.

In this paper, we use social media data to understand peoples' concerns about different variants of SARS-CoV-2 from Twitter database using two different types of analysis, Tabular Analysis and Graphical Analysis as seen from Figure 1. We found out that Omicron is currently the most popular topic (out of all other COVID Variants) with an influence score of 70.2 million. On the other hand, Alpha variant is most popular in terms of user engagements (i.e., likes and shares) with a score over 5.0 K within the measured timeframe. From the hashtag ranking, it was identified that both Omicron and Delta were the most popular COVID variants that the users were using hashtags on. It appears that for all the cases most of tweets were in English followed by French and Dutch languages. It should be mentioned that this study did not used any *Data Mining* (DM) or *Artificial Intelligence* (AI) based approaches like our previous research in global event analysis [6], [2], landslide analysis [3, 8], cardiovascular disease detection [9, 10], or person identification [11]. Rather, this paper utilizes generic statistical techniques like frequency ranking and others simplified calculations [12] to extract meaningful information which might be critical for a social scientist or decision makers who are not familiar with complex AI or DM based techniques.

2. Materials and Methods

With the help of *Application Programming Interfaces* (API), it is possible to systematically access social media posts from popular social media platforms like Twitter, Facebook, Instagram, Telegram, Snapchat etc. In this study, we used third party APIs and Interfaces like Microsoft Power Automate, Microsoft Power BI to easily access and analyze social media posts on selected keywords from Twitter. This technology architecture has been previously shown in our most recent works [2,3,6,8]. Once the social media posts are gathered using Microsoft Power Automate, they are stored in Microsoft SQL Server. Finally, Microsoft Power BI is used to display the data with *Artificial Intelligence* (AI) based tools and techniques like Regression based Key Influences, *Convolution Neural Network* (CNN) based Anomaly Detectors and Decomposition Tree Analysis as shown in [2, 3, 6, 8].

Table 1. Defining the analytical variables for social media analysis

Users	The total number of unique users who posted Tweets.
Posts	The total number of tweets downloaded.
Engagement	Total number of likes or retweets received by the post.
Influence	The sum of number of followers for each user who posted tweets on the selected keywords.
Word Cloud	A word cloud is a visual representation of textual data, which is often used to depict keyword metadata on websites, or to visualize free form text. Words within the word cloud are usually single words, and the importance of each word is shown with font size as well as color.
Hashtag Ranking	Hashtag's word or phrase preceded by a hash sign (#), used on social media websites and applications, especially Twitter, to identify digital content on a specific topic. This study ranks the hashtags extracted from twitter database on the selected keyword in descending order.
User Mentions Ranking	User Mentions (often known as @mentions) are a useful way of drawing someone's attention to a tweet message or assigning a task to them. This study ranks the user mentions extracted from twitter database on the selected keyword in descending order.
Users with Most Followers Ranking	Number of followers for the tweet user posting on the selected keyword are tracked and ranked in descending order.
Language Ranking	The tweet language is detected from the extracted tweets containing the selected keywords and then the number of detected languages are ranked in descending order.

As seen from Table 1, there are two different types of analysis conducted from twitter database. Tabular analysis included variables such as Users, Posts, Engagement, and Influence On the other hand, Graphical Analysis included Word Cloud, Hashtag Ranking, User Mentioned Ranking, User with Most Followers Ranking as well as Language Ranking. Table 1 defines these concepts in detail.

3. Results and Discussion

Tweets were extracted on 5 different keywords: COVID Variants, Omicron, Alpha Variant, Beta Variant, Gamma Variant and Delta Variant from 11th January 2022 till 20th January 2022. Along with the Tweet texts, several other variables were extracted from Twitter. These variables were Tweet language, number of users, favorite counts for each post, number of followers for each user etc. Table 2 summarizes the result. As seen from Table 2, the highest number of users posted on "COVID Variants" with more than 2 thousand users posting more than 2.1 K tweets. In terms of engagement, "Alpha Variant" received the highest attentions with likes or retweets. However, in terms of influence (i.e., the sum of number of followers for each user who posted tweets), "Omicron" was the highest. Keyword Omicron had an influence score of 70.2 M followed by 63.9 M influence score on "COVID Variant".

Table 2. Summarization of users, posts, engagement, and influence for the 6 keywords

Keywords	Users	Posts	Engagement	Influence
COVID Variants	2.0 K	2.1 K	736	63.9 M
OMICRON	1.8 K	2.1 K	506	70.2 M
ALPHA VARIANT	1.9 K	2.1 K	5.0K	62.8 M
BETA VARIANT	677	755	1.8 K	36.4 M
GAMMA VARIANT	325	351	994	27.8 M
DELTA VARIANT	1.3 K	1.4K	1.5K	24.1 M

Table 3 shows the key word frequencies with word cloud visualization as well as ranking of the popular hashtags that comes with each of the 6 keywords. For example, as seen from the word cloud of "COVID Variants", there were several key points of the discussions within Twitter, like, *unvaccinated*, *idiots* etc. These key topics refers to the current rhetoric over Anti-Vax community not getting vaccinated. In terms of most popular hashtag for "COVID Variants", "#masktup" was very popular with 141 occurrences. In terms of "Omicron", the key topics of discussions were restrictions and deaths among others, showing public concerns over government lockdowns and restrictions along with COVID related deaths. The most popular hashtags for "Alpha Variant" were "#auspol" and "#nswpol" with a cumulative frequency of 146 over the monitored period.

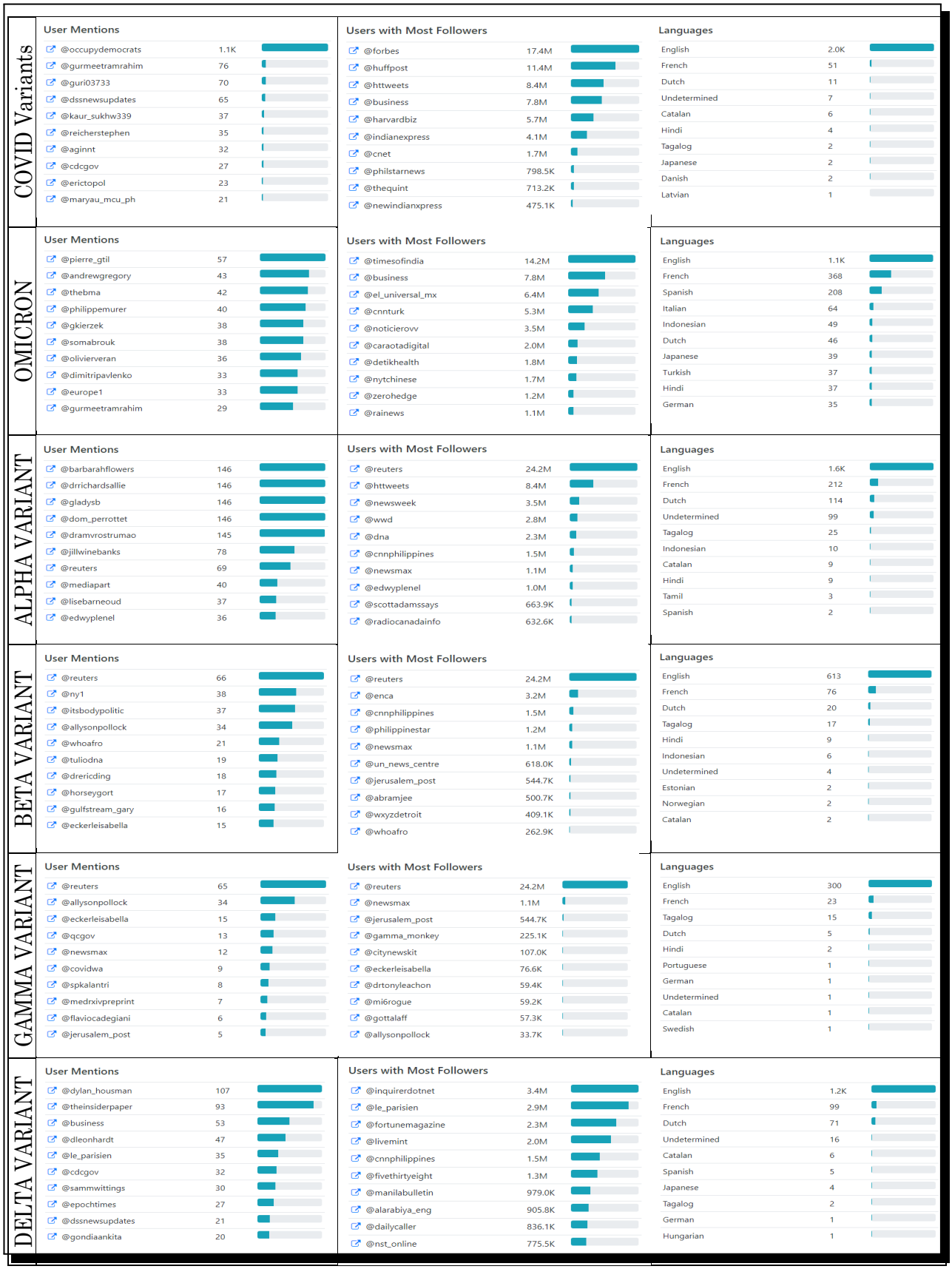
Table 4 shows user mentions (i.e., the users referred the most within tweets), users with most followers as well as the tweet language ranking. It appears that for all the cases most of tweets were in English followed by French and Dutch languages.

As seen from Table 4, @occupydemocrats is the most popular user mentioned under "COVID Variants" topic with 1.1 K mentions. @pierre_gtil was the most popular user mention for Omicron and @dylanhousman was the most popular user mention for Delta variant. The most popular sources for influences on COVID-19 Variant related posts were found to be @reuters with 24.2M, @forbes with 17.4 M, @timesofindia with 14.2 M and @inquirerdotnet with 3.4 followers.

Table 3. Word Cloud and Hashtag rankings for each of the 6 keywords

	Word Cloud	Hashtags
COVID Variants		#maskup 141 #covid19 68 #covid 36 #omicron 20 #coronavirus 10 #sarscov2 10 #delta 8 #wastebeforecase 8 #variants 7 #replenishrrf 6
OMICRON		#omicron 296 #covid19 132 #covid 58 #europe1 32 #maskup 30 #coronavirus 29 #auspol 25 #omicronvariant 19 #uptet 14 #ómicron 12
ALPHA VARIANT		#auspol 146 #nswpol 146 #omicron 48 #covid19 33 #alpha 25 #delta 17 #variant 15 #covid19wa 11 #sarscov2 10 #beta 10
BETA VARIANT		#omicron 48 #delta 43 #sarscov2 21 #variant 21 #covid19 21 #beta 19 #alpha 18 #gamma 18 #longcovid 17 #zeta 16
GAMMA VARIANT		#omicron 23 #delta 19 #sarscov2 18 #alpha 18 #beta 18 #gamma 18 #zeta 16 #variant 16 #covid19wa 11 #molnupiravir 9
DELTA VARIANT		#covid19 105 #delta 61 #omicron 45 #vaccines 31 #covid 23 #maskup 23 #omikron 21 #coronavirus 14 #us 7 #โควิด 6

Table 4. User mentions, users with most follower and language rankings for each of the keywords



	User Mentions		Users with Most Followers		Languages	
COVID Variants	@occupydemocrats	1.1K	@forbes	17.4M	English	2.0K
	@gurmeetramrahim	76	@huffpost	11.4M	French	51
	@guri03733	70	@httweets	8.4M	Dutch	11
	@dssnewsupdates	65	@business	7.8M	Undetermined	7
	@kaur_sukhw339	37	@harvardbiz	5.7M	Catalan	6
	@reicherstephen	35	@indianexpress	4.1M	Hindi	4
	@aginnt	32	@cnet	1.7M	Tagalog	2
	@cdcgov	27	@philstarnews	798.5K	Japanese	2
	@erictopol	23	@thequint	713.2K	Danish	2
	@maryau_mcu_ph	21	@newindianxpress	475.1K	Latvian	1
OMICRON	@pierre_gtil	57	@timesofindia	14.2M	English	1.1K
	@andrewgregory	43	@business	7.8M	French	368
	@thebma	42	@el_universal_mx	6.4M	Spanish	208
	@philippemurer	40	@cnnturk	5.3M	Italian	64
	@gkierzek	38	@noticierovv	3.5M	Indonesian	49
	@somabrouk	38	@caraotadigital	2.0M	Dutch	46
	@olivierveran	36	@detikhealth	1.8M	Japanese	39
	@dimitripavlenko	33	@nytchinese	1.7M	Turkish	37
	@europe1	33	@zerohedge	1.2M	Hindi	37
	@gurmeetramrahim	29	@rainews	1.1M	German	35
ALPHA VARIANT	@barbarahflowers	146	@reuters	24.2M	English	1.6K
	@drrichardsallie	146	@httweets	8.4M	French	212
	@gladysb	146	@newsweek	3.5M	Dutch	114
	@dom_perrottet	146	@wwd	2.8M	Undetermined	99
	@dramvrostrumao	145	@dna	2.3M	Tagalog	25
	@jillwinebanks	78	@cnnphilippines	1.5M	Indonesian	10
	@reuters	69	@newsmax	1.1M	Catalan	9
	@mediapart	40	@edwyplenel	1.0M	Hindi	9
	@lisebarneoud	37	@scottadamssays	663.9K	Tamil	3
	@edwyplenel	36	@radiocanadainfo	632.6K	Spanish	2
BETA VARIANT	@reuters	66	@reuters	24.2M	English	613
	@ny1	38	@enca	3.2M	French	76
	@itsbodypolitic	37	@cnnphilippines	1.5M	Dutch	20
	@allysonpollock	34	@philippinestar	1.2M	Tagalog	17
	@whoafro	21	@newsmax	1.1M	Hindi	9
	@tuliodna	19	@un_news_centre	618.0K	Indonesian	6
	@drericding	18	@jerusalem_post	544.7K	Undetermined	4
	@horseygort	17	@abramjee	500.7K	Estonian	2
	@gulfstream_gary	16	@wxyzdetroit	409.1K	Norwegian	2
	@eckerleisabella	15	@whoafro	262.9K	Catalan	2
GAMMA VARIANT	@reuters	65	@reuters	24.2M	English	300
	@allysonpollock	34	@newsmax	1.1M	French	23
	@eckerleisabella	15	@jerusalem_post	544.7K	Tagalog	15
	@qcgov	13	@gamma_monkey	225.1K	Dutch	5
	@newsmax	12	@citynewskit	107.0K	Hindi	2
	@covidwa	9	@eckerleisabella	76.6K	Portuguese	1
	@spkalantri	8	@drtonyleachon	59.4K	German	1
	@medrxivpreprint	7	@mi6rogue	59.2K	Undetermined	1
	@flaviocadegiani	6	@gottalaff	57.3K	Catalan	1
	@jerusalem_post	5	@allysonpollock	33.7K	Swedish	1
DELTA VARIANT	@dylan_housman	107	@inquirerdotnet	3.4M	English	1.2K
	@theinsiderpaper	93	@le_parisien	2.9M	French	99
	@business	53	@fortunemagazine	2.3M	Dutch	71
	@dleonhardt	47	@livemint	2.0M	Undetermined	16
	@le_parisien	35	@cnnphilippines	1.5M	Catalan	6
	@cdcgov	32	@fivethirtyeight	1.3M	Spanish	5
	@sammwittings	30	@manilabulletin	979.0K	Japanese	4
	@epochtimes	27	@alarabiya_eng	905.8K	Tagalog	2
	@dssnewsupdates	21	@dailycaller	836.1K	German	1
	@gondiaankita	20	@nst_online	775.5K	Hungarian	1

4. Conclusion

The most popular variants that people talked about during the monitored period were "Omicron" and "Delta" as seen from hashtag rankings of "COVID Variants" in Table 3. This is mainly because both these variants display aggressive behaviors with delta being more deadly and omicron being more transmissible. Using techniques demonstrated in this study, significant insights on COVID-19 variants or any other topics could be successfully extracted from Social Media sources like Twitter, Facebook, Instagram, LinkedIn, Snapchat, Telegram, or others.

This study did not utilize any *Artificial Intelligence* (AI) based approach in analyzing COVID-19 data. In future, several AI infused techniques on COVID-19 data may result in more interesting results. It should be mentioned that we have already started, working on COVID-19 data with AI infused techniques like sentiment analysis and *named entity recognition* (NER) to intelligently extract meaningful critical information [4, 7]. These AI based techniques (i.e., Sentiment Analysis and NER) has also been successfully used in analyzing global breaking events in [2, 6] as well as [5].

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] Datareportal, Digital Around the World, 2021, available at: https://datareportal.com/global-digital-overview, accessed January 12, 2022.

[2] F. Sufi, AI-GlobalEvents: A software for analyzing, identifying and explaining global events with Artificial Intelligence, *Software Impacts* **11**(100218) (2022), 1 – 5, DOI: 10.1016/j.simpa.2022.100218.

[3] F.K. Sufi, AI-Landslide: Software for acquiring hidden insights from global landslide data using Artificial Intelligence, *Software Impacts* **10** (100177) (2022), 1 – 5, DOI: 10.1016/j.simpa.2021.100177.

[4] F.K. Sufi, AI-based automated extraction of entities, entity categories and sentiments on COVID-19 situation, *IEEE Dataport* (2021), DOI: 10.21227/sawp-ax73.

[5] F.K. Sufi, Analyzing natural disaster related tweets with AI and NLP based services, *IEEE Dataport* (2021), DOI: 10.21227/p0ed-cb23.

[6] F.K. Sufi and M. Alsulami, Automated multidimensional analysis of global events with entity detection, sentiment analysis and anomaly detection, *IEEE Access* **9** (2021), 52449 – 152460, DOI: 10.1109/ACCESS.2021.3127571.

[7] F.K. Sufi and M. Alsulami, AI-based automated extraction of location-oriented COVID-19 sentiments, *Computers, Materials & Continua* **72**(2) (2022), 3631 – 3649, DOI: 10.32604/cmc.2022.026272.

[8] F.K. Sufi and M. Alsulami, Knowledge discovery of global landslides using automated machine learning algorithms, *IEEE Access* **9** (2021), 131400 – 131419, DOI: 10.1109/ACCESS.2021.3115043.

[9] F. Sufi and I. Khalil, Diagnosis of cardiovascular abnormalities from compressed ECG: a data mining-based approach, *IEEE Transactions on Information Technology in Biomedicine* **15**(1) (2010), 33 – 39, DOI: 10.1109/TITB.2010.2094197.

[10] F. Sufi and I. Khalil, A clustering based system for instant detection of cardiac abnormalities from compressed ECG, *Expert Systems with Applications* **38**(5) (2011), 4705 – 4713, DOI: 10.1016/j.eswa.2010.08.149.

[11] F. Sufi and I. Khalil, Faster person identification using compressed ECG in time critical wireless telecardiology applications, *Journal of Network and Computer Applications* **34**(1) (2011), 282 – 293, DOI: 10.1016/j.jnca.2010.07.004.

[12] F. Sufi, Q. Fang, I. Khalil and S.S Mahmoud, Novel methods of faster cardiovascular diagnosis in wireless telecardiology, *IEEE Journal on Selected Areas in Communications* **27**(4) (2009), 537 – 552, DOI: 10.1109/JSAC.2009.090515.

[13] World Health Organization, Coronavirus disease (COVID-19) pandemic, 2022, available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019, accessed January 15, 2022.

[14] World Health Organization, Tracking SARS-CoV-2 variants, 2021, available at: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/, accessed January 15, 2022.