Extracting factors associated with vaccination from Twitter data and mapping to behavioral models

Md. Rafiul Biswas, Zubair Shah

Taylor & Francis
Taylor & Francis Group

Human Vaccines & Immunotherapeutics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/khvi20

Extracting factors associated with vaccination from Twitter data and mapping to behavioral models

Md. Rafiul Biswas & Zubair Shah

To cite this article: Md. Rafiul Biswas & Zubair Shah (2023) Extracting factors associated with vaccination from Twitter data and mapping to behavioral models, Human Vaccines & Immunotherapeutics, 19:3, 2281729, DOI: 10.1080/21645515.2023.2281729

To link to this article: https://doi.org/10.1080/21645515.2023.2281729

HUMAN VACCINES & IMMUNOTHERAPEUTICS
2023, VOL. 19, NO. 3, 2281729
https://doi.org/10.1080/21645515.2023.2281729

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS | Check for updates

Extracting factors associated with vaccination from Twitter data and mapping to behavioral models

Md. Rafiul Biswas and Zubair Shah

Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

ABSTRACT

Social media platform, particularly Twitter, is a rich data source that allows monitoring of public opinions and attitudes toward vaccines.Established behavioral models like the 5C psychological antecedents model and the Health Belief Model (HBM) provide a well-structured framework for analyzing shifts in vaccine-related behavior. This study examines if the extracted data from Twitter contains valuable insights regarding public attitudes toward vaccines and can be mapped to two behavioral models. This study focuses on the Arab population, and a search was carried out on Twitter using: ’ تلقيحي OR تطعيم OR تطعيمات OR لقاح OR لقاحات’ for two years from January 2020 to January 2022. Then, BERTopicmodeling was applied, and several topics were extracted. Finally, the topics were manually mapped to the factors of the 5C model and HBM. 1,068,466 unique users posted 3,368,258 vaccine-related tweets in Arabic. Topic modeling generated 25 topics, which were mapped to the 15 factors of the 5C model and HBM. Among the users, 32.87%were male, and 18.06% were female. A significant 55.77% of the users were from the MENA (Middle East and North Africa) region. Twitter users were more inclined to accept vaccines when they trusted vaccine safety and effectiveness, but vaccine hesitancy increased due to conspiracy theories and misinformation. The association of topics with these theoretical frameworks reveals the availability and diversity of Twitter data that can predict behavioral change toward vaccines. It allows the preparation of timely and effective interventions for vaccination programs compared to traditional methods.

ARTICLE HISTORY
Received 7 March 2023
Revised 26 October 2023
Accepted 5 November 2023

KEYWORDS
COVID-19; vaccination; behavior; attitudes; 5C model; health belief model

Introduction

Background

Public attitudes toward disease prevention programs, specifically vaccines, have been a topic of great importance in public health since the coronavirus pandemic has impacted all facets of life. Over the years, there has been a noticeable shift in how individuals perceive and prioritize vaccines as preventive measures for controlling diseases.[1] Vaccines are an important tool in public health to prevent infections and/or severe health outcomes.[2] Despite evidence of the safety and effectiveness of vaccines, misperceptions about safety and effectiveness persist. While there has been a positive change in public attitudes toward disease prevention programs like vaccines, it is essential to acknowledge the existing challenges like vaccine hesitancy and vaccine refusal. Studies[3–5] suggest that negative vaccine information from news media, health practitioners, and celebrities can increase vaccine refusal and hesitancy and affect vaccine uptake and coverage. A low vaccination coverage will increase the risk of infection and the severity of outbreaks.[6–8] For instance, cervical cancer is caused by Human Papillomavirus (HPV) infections, and HPV vaccines have been shown to reduce the prevalence of HPV. However, acceptance of the HPV vaccine is the leading worldwide barrier to HPV vaccination coverage.[9–11] Understanding the factors related to vaccine attitudes bears substantial importance because it helps identify the root causes contributing to vaccine hesitancy. This understanding will enable the policymaker to plan more effectively to counter misinformation, build trust, and ultimately curb vaccine hesitancy. Researchers have developed various theoretical models such as Health Belief Model (HBM),[12] Theory of Planned Behavior (TPB),[13] 5C Model,[14] Social Cognitive Theory (SCT),[15] and others for health promotion and disease prevention to conceptualize the context of behavioral change. Each model offers unique perspectives and insights into behavior change processes, and practitioners often select or combine models based on their interventions' specific context and goals. Studies[16,17] used the 5C model to describe vaccine behavior factors to understand COVID-19 vaccine hesitancy and acceptance. Studies[18–20] used HBM to find factors that affect attitudes toward COVID-19 vaccinations. The HBM has also been used to identify beliefs and attitudes toward seasonal influenza[21] and swine flu vaccines.[22] Some of the studies[23–25] used combinedly TPB, HBM, and 5C models to prepare questionnaires for surveys and interviews to predict public behavior and attitudes toward vaccines. The usage of these models by researchers in vaccine behavior prediction shows the acceptability of the model. There are differences in the framing of the model structures. Our goal is not to find a comparison among behavioral models. Rather, we aim to

CONTACT Zubair Shah zshah@hbku.edu.qa Division of ICT, College of Science and Engineering, Hamad Bin Khalifa University (HBKU), A101-F, LAS Building, Education City, PO Box 34110, Doha, Qatar.

Supplemental data for this article can be accessed on the publisher's website at https://doi.org/10.1080/21645515.2023.2281729.

validate the quality of Twitter data with any framework of the behavioral model. That's why we have chosen two models: HBM and 5C model. The 5C model has five constructs: confidence, complacency, constraints, calculation, and collective responsibility. Confidence refers to trust in the safety and effectiveness of the vaccine, the delivery system, and the policymaker's role.[26] Complacency relates to how individuals perceive disease threats and deem the necessity of vaccination. Constraints discuss the physical and psychological barriers to vaccination. Calculation refers to the risk of getting an infection and the benefits of vaccination. Similarly, the HBM also has five constructs: perceived susceptibility, perceived severity, perceived benefits, perceived barriers, and cues to action.[21,27] Perceived susceptibility refers to the belief in the potential risk of contracting the disease. Perceived benefit refers to the belief that the vaccine will reduce disease threats. Perceived barriers refer to lower intention to vaccination due to physical, financial, and psychological factors. Perceived severity refers to the belief of severe consequences/losses caused by diseases due to infection.

In our previous study, we prepared questionnaires by following the HBM and performed a survey in the Arab world to find the public attitudes and behavior toward the COVID-19 vaccine.[28] Our survey study found that the majority of participants were optimistic about vaccine acceptance, though concerns about the rapid development of vaccines and potential long-term side effects led some to prefer delaying vaccination. Furthermore, variations in acceptance were observed based on demographics and community norms. The current study was designed to see if vaccine attitudes in the MENA region can be gauged using Twitter data, eliminating the need to develop or employ questionnaires. This approach is premised on the hypothesis that the concerns that surfaced in the survey study would be reflected in the Twitter data during the survey period, providing a complimentary source for understanding vaccine-related concerns discussed on social media. To obtain the goal, we considered the MENA region as our experimental region, and Arabic vaccine-related tweets posted between January 2020 to January 2022 were collected and used as study data. We chose this period because the COVID-19 pandemic lasted, and the vaccine discussions reached a peak on Twitter.[29]

The proposed method acts as a social media information grading tool that can produce localized estimates of the quality of information people share and is exposed to via social media and their vaccine attitudes. The location analysis of the Twitter data reflects that the target population belongs to MENA region countries. Analysis data from the Twitter platform forms costs less to administer, collects data faster, and is broader than a traditional questionnaire-based survey, which is resource-intensive, relatively slow to report, and may not reach inaccessible sub-populations.[30] Also, survey requires developing a scale to measure general behaviors, attitudes, and hypothetical scenarios that cannot be captured in a single variable or item.[31,32] Using Twitter data for surveillance provides ongoing, real-time information, complementing insights from periodic surveys. In addition, observing social media is unobtrusive and avoids recall bias[33] and the Hawthorne effect.[34]

Vaccine communication in social media

Researchers usually collect public health data through traditional registries, surveying people, and public health reports. During the COVID-19 pandemic, surveys became one of the most widely used methods to identify vaccine hesitancy and acceptance. Biswas MR[35] listed 82 studies, and Sallam M[36] listed 30 studies where authors performed surveys to measure COVID-19 vaccine hesitancy. Aw J[37] listed 97 studies, 87 of which were online surveys and the rest cross-sectional studies (e.g., telephone interviews, paper questionnaires, and group discussions). A systemic review reported 22 studies between 2007 and 2017 that performed surveys to measure parental behavior toward vaccinating their children.[38]

Social media has been a source of information during pandemics for decades.[39] A growing corpus of literature has employed social media sites such as Twitter and Facebook to find out public perceptions toward vaccines.[40] Social media might influence vaccination decisions by delivering information on the perceived personal risk of vaccine-preventable diseases or vaccine side effects. The simulated Twitter posts employed through a survey study showed how anticipated regret and consequences can significantly influence vaccination intentions.[41] Negative news circulated through print media such as newspaper, radio, and television becomes dominant over positive news and influence a large number of the population toward vaccine decision.[42] The Twitter platform can be monitored to extract and analyze signal indicators of population-level health outcomes.[43–47] Examples that are validated against outcome data include infectious diseases like influenza and cholera,[48,49] HPV coverage,[50] and heart disease.[51] Also, Twitter data has been used as a source for mining public opinions in different countries during the COVID-19 pandemic.[52]

There is a gap in analyzing public opinion from Twitter data in the Arab world. Only a few studies worked on Twitter data in Arabic to extract sentiment and attitudes. There are few datasets[53–57] publicly available related to COVID-19 in the Arabic language. There is a lack of a natural language toolkit for the Arabic language for data preprocessing, sentiment analysis, and topic modeling. While Mubarak H[58] collected vaccine-related Arabic tweets and authors[52,59,60] analyzed sentiments toward vaccines from Arabic tweets, these studies lack identifying vaccine hesitancy and vaccine uptake.

Accessing Twitter data is easy and more practical to perform real-time analysis of public sentiment and opinion on COVID-19 vaccines. Thus, it would be interesting to observe which topics influence public opinions and what are the predictors of these topics to achieve public attraction. It is crucial for the policymaker to understand citizen and resident attitudes aiding decision-making and future planning. COVID-19 vaccine hesitancy was dynamic because public sentiment and emotions changed with time, and the situation needs real-time analysis to identify changes in public attitudes. By automating the process, it is possible to collect Twitter data and analyze it in less than one month.

This study primarily illustrates the innovative methodology of extracting and analyzing public attitudes toward vaccines from Twitter data, utilizing the frameworks of the Health

Belief Model (HBM) and the 5C model. The emphasis is placed on exploring the prevalence of vaccine-related content on Twitter to gain insights into prevailing public attitudes and perceptions. The association of Twitter topics with these factors enables a comprehensive understanding of the available data and its significance in predicting behavioral shifts related to vaccine acceptance or hesitancy. By exploring this innovative method, the study showcased the potential to gain valuable insights from analyzing Twitter data for understanding public vaccine attitudes, offering an alternative to conventional approaches.

(i) What are the factors associated with public attitudes toward vaccines?
(ii) How do the extracted factors map in the behavioral models of the HBM and 5C model?

Method

This research was carried out in several steps, such as data preprocessing, topic modeling, and mapping the identified topics to the behavioral model. According to the terms and conditions of the Twitter Academic Research API, Twitter data was only used for research purposes and has not been shared outside the research group. Since this research does not reveal any personal information, it did not require a review by an institutional review board. All the authors agreed on the process, and there was no conflict of interest between the researchers.

Data collection and preprocessing

Vaccine-related Arabic tweets were downloaded using the Twitter Academic Research Application Programming Interface (API)[61] from January 2020 to January 2022. The search terms used to collect Arabic vaccine-related tweets consisted of five keywords: ‘تلقيحي OR تطعيم OR تطعيمات OR لقاح OR لقاحات’ (English translation: ‘vaccine OR vaccination OR immunization OR vax OR vaccines). The search terms were validated by four native Arabic speakers from Egypt, Yemen, Qatar, and Syria. The keywords contain the main form of the vaccine word and can be used with any suffixes and prefixes. Tweets were stored in PostgreSQL database tables (i.e., users table, tweet table). Tweet ID was the primary key used to identify each tweet uniquely and avoid duplication of data entry. The users' table consisted of 21 columns, among which user_id, name, username, location, verified, followers_count, following_count, tweet_count columns are used. The tweets table consisted of 28 columns: text, reply_count, retweets_count, like_count, etc.

Twitter data is noisy, so it needs to be preprocessed to prepare for analysis. First, non-Arabic tweets were identified using a language field and removed. Next, retweets were determined by the "RT @" or "@ RT" or "RT" string and removed. Next, non-printable characters such as emojis and images, punctuation marks, and Unified Resource Links (URLs) were removed. Afterward, Arabic stopwords (i.e., very common words in a sentence with less meaning) were identified using a GitHub repository[62] and removed from the analysis. Because stopwords do not add much meaning to sentences, they can be ignored during natural language processing. Next, usernames were normalized by replacing username mentions ("@username" in the tweets) with empty strings to maintain anonymity. Word tokenization was performed using the NLTK library[63] to divide a large quantity of text into smaller parts called tokens. Finally, the data were prepared to train in a machine learning model. The data preprocessing flow is illustrated in Figure 1.

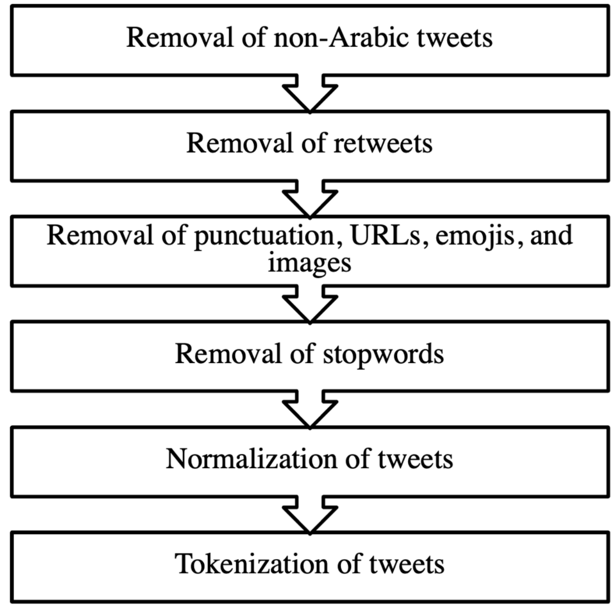

Figure 1. Data preprocessing.

Finding user demographics

We identified two types of user demographics, such as gender and user location. Gender reveals important characteristics of attitudes in health-related issues.[57] We used the name entity column from the Twitter dataset to identify gender. A GitHub dataset[64] contains 6000 most popular names in Arabic labeled with gender (male and female). We used this dataset as a benchmark. We developed a machine model using a random forest classifier to predict the gender of an individual. The model considers names to have at least two characters. Any name containing ابو is considered male and whereas any name containing عبد is considered female. Some names were in English, and we used the widely used Name Entity Recognition (NER) model to detect gender.[65] Many users didn't provide complete names rather they used symbols and became ambiguous for the model to predict gender. We assigned the label 'unknown' for this type of gender.

User geo location possesses indispensable information regarding disease hot-spot prediction, epidemic spread monitoring, and risk mapping.[66] Many users provide location while creating their profiles. Users can also write any items in their location, such as country name, state, and city granularities. So, we developed Python code to predict country names from user-specific locations at national and sub-national levels. Some users didn't write country names. They write only the city name or state name. We extracted country names from city granularities or states. Country names can also be extracted from other column country codes.

Topic modeling

Topic modeling is an unsupervised machine-learning technique that organizes a large volume of documents into several small clusters representing different constructs.[67] This study applied BERTopic modeling to extract topics from the tweets.[68] BERTopic modeling embedded with a pre-trained transformer model (e.g., HDBSCAN) and a c-TF-IDF method[68] can generate easily interpretable topics. BERTopic modeling was chosen over other topic modeling techniques because it uses a c-TF-IDF method to keep track of changes in a corpus containing short messages (e.g., Twitter).[69] The default setting for the BERTopic is Language: 'English' and embedding model: 'all-MiniLM-L6-v2'. BERTopic supports over 50+ languages by writing 'multilingual'. Multilingualism was selected because most tweets are in Arabic; however, some are mixed with Arabic and other languages. In this scenario, the multilingual model performs better than the monolingual model.[70] The default embedding model has large parameters and needs high computational power for large datasets. We choose 'paraphrase-multilingual-MiniLM-L12-v2' embedding model, which performs faster and better with less computational power. The model was trained in Google Colab Pro+ with Python 3.7 environments. The total number of tweets was about 3.36 million. The RAM was insufficient to train all the data at one time, so the data were sorted according to the date column metadata and split into six groups with a size of 580,000 tweets. Then, we applied the BERTopic topic modeling algorithm. There is a need to set some parameters such as minimum topic size, calculation of probabilities, model name, and others to train the model. BERTopic requires a minimum input topic size to train the data. Based on the input topics number and dataset size, it generates topics. If the minimum input is less, it generates more topics. We started with a small input size of 20, then increased gradually and analyzed the topics. A larger number of topics are more similar and often need to combine to avoid duplicate topics and result in a meaningful topic. We stopped with the input size of 200 due to computational power limitations. The 'calculate_probabilities' was then assigned to true to calculate the probabilities of all topics across all documents instead of only the assigned topic. This, however, slows down computation and may increase memory usage.[68]

BERTopic generated a total of 540 topics from the six training sessions. As we split the data into six sets, some topics were redundant. The first author identified the redundant topics manually and removed them from the analysis. The second author validated the work done by the first author. Similar topics that contained identical discussions were manually merged into one topic. Then, keywords from topic modeling outputs were identified. A string-matching search was applied to complete datasets to determine the presence of the selected topic in the tweets. Logical operators (i.e., '|,' '&') were used to make search strings. The topics were labeled based on the tweets' highest percentages of associated words. Topic names were then determined through discussion. The topics were manually mapped to each factor of the behavioral model. The keywords and tweets of each topic were analyzed to find the similarities to fit into each factor of the 5C model and HBM. This process was carried out through discussion and agreement with the authors. Prominent topics that got the highest number of likes and retweets were discovered. Then, the mean number of likes, retweets, and replies for each topic was calculated. Last, the interaction rate for each topic was calculated by adding the total number of likes, retweets, and replies per topic and dividing it by the total number of followers per topic. These analyses provided further insights into the most prominent topics and public interest.

Constructs of theory

Researchers use the 5C Model[14,16,17,25,71–74] and HBM[18–21,24,25] to develop questionnaires and survey people to measure public behaviors and attitudes toward vaccines. The 5C model provides a theoretical framework for understanding the influencing factors of public attitudes that contribute to vaccine hesitancy.[25] The 5C model has five constructs (i.e., categories), and each construct has several factors (i.e., subcategories) (see Table 1). In this study, we applied a positive mark (+) to the factors that positively influenced individuals' beliefs toward vaccination (i.e., vaccine acceptant) and a negative mark to the factors that negatively (-) influenced their beliefs toward vaccination (i.e., vaccine-hesitant).[14] Confidence has five factors: i) vaccination attitudes toward vaccines (+) ii) Beliefs about medicine: benefits (+) iii) Beliefs about medicine: harms (-) iv) Trust in the role of public authorities (+) v) Conspiracy mentality (-). Vaccine effectiveness and the role of public authorities are positively correlated with vaccine belief. Harms of medicines and conspiracy mentality are negatively correlated with vaccine belief. Similarly, "perceived threat due to infectious diseases (+)" intensifies the need for a vaccine and is positively correlated to vaccination belief. "Risk attitudes of vaccine" and "consideration of future consequences" are negatively correlated with vaccine decisions, indicating hesitancy toward preventive measures. We have followed the scales[14] to construct the 5C model (for more detail, refer to supplementary file S1).

HBM has been extensively used as a conceptual framework to evaluate and predict vaccine behavior and

```
topic_model = BERTopic(model='paraphrase-multilingual-MiniLM-L12-v2',
language="multilingual", calculate_probabilities=True, min_topic_size=200,
verbose=True, low_memory= True, n_gram_range=(1,3))
topics, probs = topic_model.fit_transform(docs)
```

Table 1. Constructs and factors from existing 5C model and HBM.

Model	Constructs	Factors
5C model	Confidence	Attitudes toward vaccines (+)
		Beliefs about medicine: benefits (+)
		Beliefs about medicine: harms (-)
		Trust in the role of public authorities (+)
		Conspiracy mentality (-)
	Complacency	Perceived threat due to infectious diseases (+)
		Consideration of future consequences (-)
		Perceived risk of vaccination (-)
	Constraints	Affordability and willingness-to-pay (+)
		Geographical Accessibility (+)
		Perceived time pressure (-)
	Calculation	Influential factors in vaccine decision (+)
		Risk calculation of vaccine (-)
	Collective responsibility	Collectivism (+)
		Communal Orientation (+)
HBM	Perceived susceptibility	Perceived risk of infection (+)
		Perceived risk of vaccine (-)
	Perceived severity	Serious consequences of Coronavirus (-)
		Serious complications of vaccine (-)
		Social and financial consequences
	Perceived benefits	Perceived vaccine effectiveness (+)
		Decrease in infection rate (+)
		Reduction of complications (+)
	Perceived barriers	Access to vaccination centers (+)
		Perception of vaccine side effects (-)
		Personal or family experience with vaccination (-)
		Introduction to a new vaccine (-)
	Cues to action	Media recommended vaccines (+)
		Influential leader recommended vaccine (+)
		Government Recommended vaccines (+)

(+) hypothesized positive attitudes toward vaccination, (-) hypothesized negative attitudes toward vaccination.

attitudes.[18–21,24,25,75] Some studies used five beliefs, whereas some studies used six beliefs for HBM construction. In this study, five beliefs (see Table 1) have been used and derived from studies.[19,22,25,27] The HBM model has five constructs (i.e., categories), and each construct has several factors (i.e., subcategories). For example, the construct "Perceived susceptibility" has two factors: "Risk of infection" and "Risk of vaccine". "Risk of infection" is marked as positive (+) because it positively influences vaccine acceptance. This means that when individuals perceive themselves as at risk of contracting a disease, they are more likely to accept vaccination as a preventive measure. On the other hand, "Risk of vaccine" is marked as negative (-) because it negatively affects individuals' beliefs and attitudes toward vaccines. When people have worries or doubts about the safety or efficacy of vaccines, they may be hesitant to accept vaccination. The detailed construction of HBM has been shown in supplementary file S1.

Mapping Twitter topics to the factors of 5C model and HBM

The identified topics are to be mapped to the factors of the 5C model and HBM. There are three ways that a topic can be mapped to a factor. Firstly, one topic is directly mapped to one factor. For instance, the topic "willingness to accept or refuse vaccine" can be mapped to the factor "attitudes toward vaccine." Secondly, each factor covers significant aspects of public behavior, and one topic may not be sufficient to interpret a factor completely. Combining two or more topics is sometimes necessary to fully interpret a factor. For instance, two topics, "trust in the role of healthcare professionals" and "trust in the role of vaccine manufacturer" can be mapped to the factor "trust in the role of public authorities." However, each topic possesses cumulative information and can be mapped into two different factors. In this study, we took care not to duplicate topics when mapping them to factors to ensure clarity and avoid confusion. By recognizing the interplay between topics and factors, a more meaningful and comprehensive analysis of public discussions and behavior regarding vaccines can be achieved. Figure 2 shows the mapping of the topic to the factor.

Finding user engagements with Twitter topics

We distinguished prominent topics that got the highest number of likes and retweets. We calculated the mean number of likes, retweets, and replies for each topic. Then, we calculated the interaction rate for each topic by the following Equation (1) by summing the total number of likes, retweets, and replies count per topic divided by the sum of the total number of followers per topic. This analysis provided further insights into the most prominent topics and public interest in it.

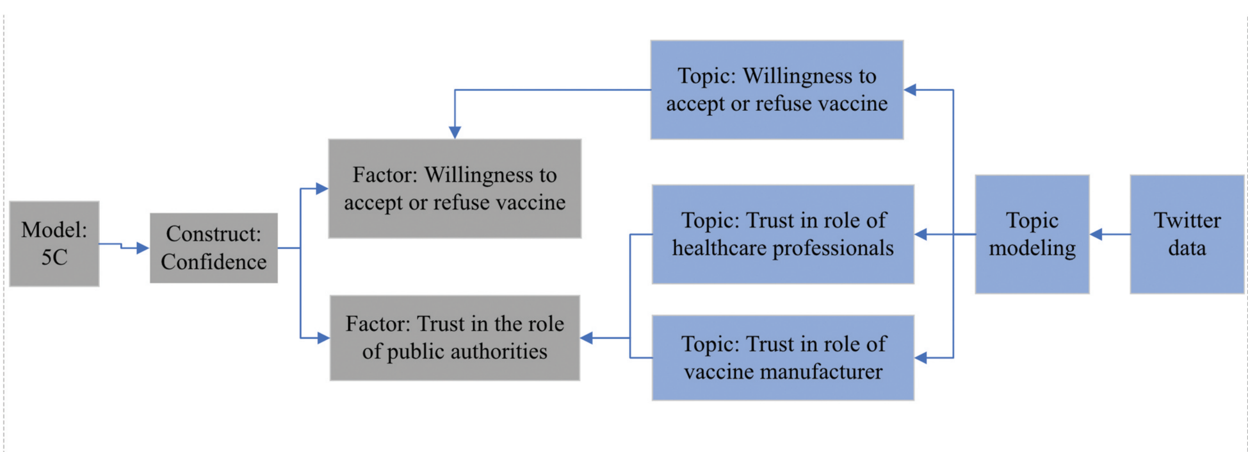

Figure 2. Mapping topics to factors.

$$interaction_{rate} = \frac{likes_{count} + retweets_{count} + replies_{count}}{followers_{count}} \quad (1)$$

Results

Study data

A total of 9,320,630 vaccine-related tweets were downloaded from 01 January 2020 to 31 January 2022 (762 days). Of these 82,396 duplicate tweets were removed. Then, 5,849,772 retweets were eliminated, and 3,388,462 unique tweets remained. Next 20,204 tweets posted in a language other than Arabic were removed, leaving 3,368,258 unique tweets posted in the Arabic language. Figure 3 shows the results of the data-cleaning process.

Figure 4 shows the weekly vaccine discussions on the Twitter platform between January 2020 and January 2022. From January 2020 to mid-March 2020, the range of tweets about vaccines was meager ($n < 10000$). A sudden vaccine discussion increase ($n = 14993$ to 34,568) occurred from mid-March 2020 to the end of March 2020.[76] At that time, the coronavirus spread across most of the countries in the world.[30] With the release of trial vaccines between November 2020 and December 2020, COVID-19 vaccine-related tweets spiked ($n = 38{,}422$ to 72,312). Twitter volume peaked ($n = 72{,}312$) in December 2020 when the Pfizer/BioNTech vaccine was officially approved.[77] After that period, vaccine discussions decreased but remained a trending topic in Arab countries. During mid-October to mid-December of 2021, vaccine tweets decreased significantly ($n = 7567$ to $n = 14008$). It may happen due to the organizing of first-time FIFA Arab Cup in Qatar which was held as a preparation for the mega event FIFA World Cup Qatar 2022. Arab Twitter users were more excited FIFA World Cup and tweeted about it. However, there might be other factors that could cause a decline in the curve. A sharp spike ($n = 36{,}521$) was again observed in January 2022 when some Arab countries started administering a third dose (booster dose) of the COVID-19 vaccine.

User demographics

Among the 1,069,229 unique Twitter users in the study, 351,451 (32.87%) were male, 193,103 (18.06%) were female, and 524,670 (49.07%) had an unknown gender (see Figure 5). Random forest classifier model predicted gender with an accuracy score of 0.85, F1 score of 0.90, precision score of 0.88, and recall score of 0.92. The higher percentages (49.07%) of unknown because many Twitter users used symbols, punctuation, remarks, and meaningless words. Our model was only capable of predicting gender in English and Arabic names. Some used different languages, such as Urdu, Chinese, French, Spanish, and other languages, which were predicted as unknown.

Among the 1,069,229 unique Twitter users analyzed, 203,590 users (19%) provided their location information. Out of these, 113,550 users (55.77% of 203,590) were found to be from the MENA (Middle East and North Africa) region. The country with the highest number of Twitter users in the study is Saudi Arabia, with 58,361 individuals. Egypt follows closely behind with 21,512 users. Kuwait ranks third with 9,636 Twitter users. Other countries in the MENA region and their corresponding Twitter user counts are as follows: UAE (3,594), Lebanon (4,101), Oman (3,140), Algeria (1,602), Jordan (2,521), Iraq (4,133), Bahrain (1,274), Qatar (1,531), Yemen (1,523), Iran (1,267), and Libya (2,468). The number of users per country in the MENA region is shown in Figure 6. Many

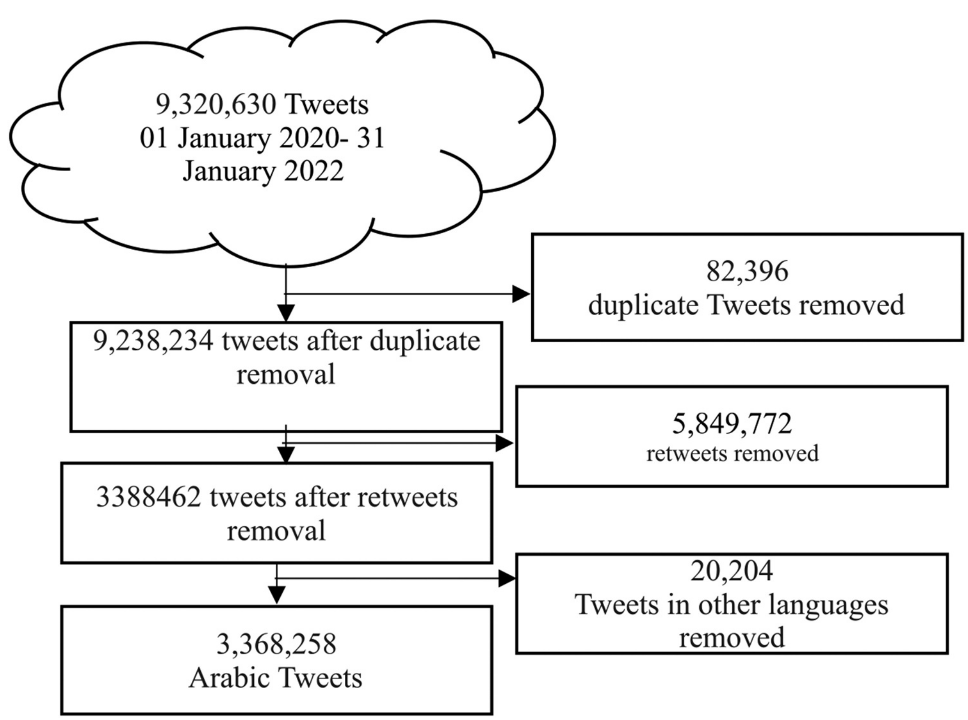

Figure 3. Flowchart of data cleaning.

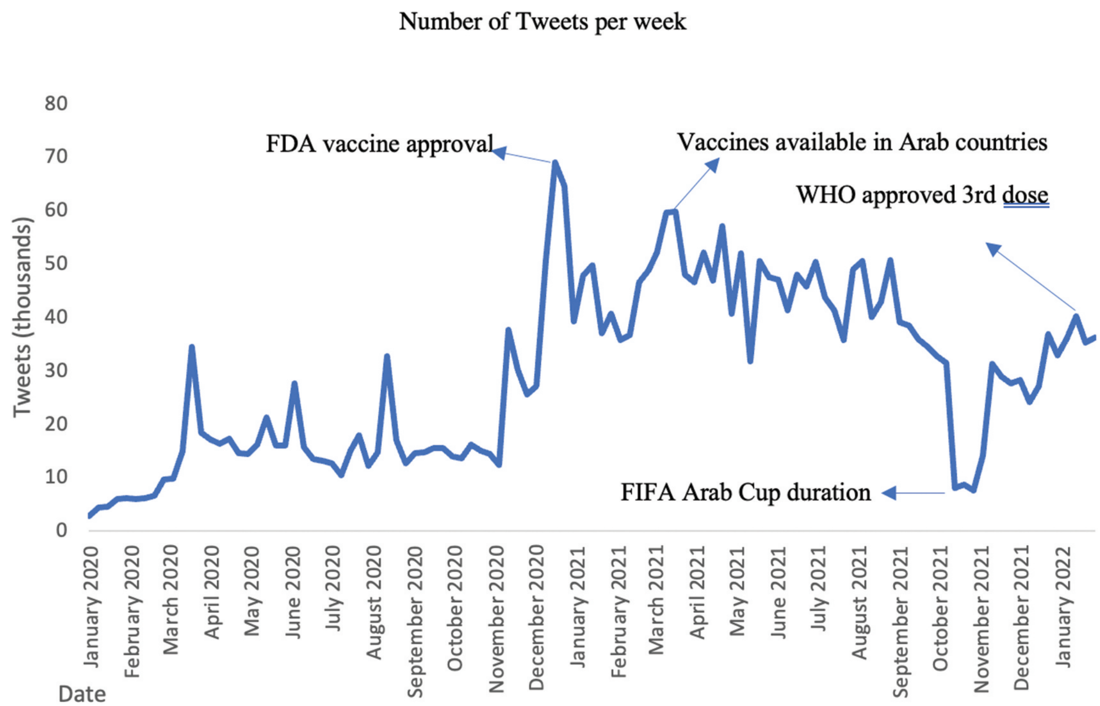

Figure 4. Weekly vaccine discussions on Twitter platform.

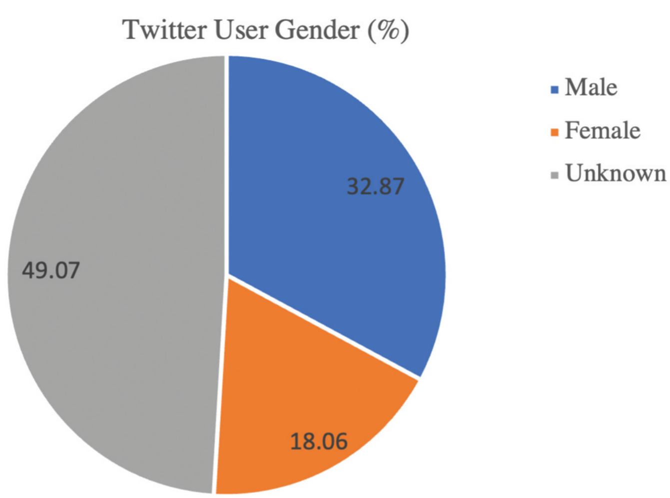

Figure 5. Gender classification in Twitter user.

users did not share their location. Instead, they wrote meaningless symbols, dots, and sentences. Some users wrote several locations along with MENA countries. Other countries were reported as Canada, the UK, the USA, Spain, Australia, and France. It is important to note that location identification on Twitter may not always be accurate or up-to-date. For example, a user may mention being located in the UAE but could currently be residing in the USA. Therefore, the location information provided on Twitter should be interpreted with caution, as it may not always reflect the true or current whereabouts of the users.

Topics extracted from tweets

The BERTopic modeling algorithm generated 540 topics through six iterations. Many topics are identical, so only unique topics were considered. Some topics were close in meaning, so they were merged. Finally, 29 topics were finalized (see Table 2). Each topic was labeled based on frequently used words and relevant tweets. Some topics could be represented as both positive and negative and marked as (±), topics related to hesitancy/barriers to vaccination were marked as negative (-), and topics related to positive attitude/intention to vaccination were marked as positive (+). Tweets were searched for relevant keywords to find the percentages of tweets per topic and shown as PT* - Percentages of Topic. The last column of Table 2 shows the translation of original Arabic tweets to English. More details about each topic and its association with the 5C model and HBM are described in the following subsections.

Mapped Twitter topics to the 5C model and HBM

Topics identified from tweets describe public attitudes and behaviors toward vaccines. The identified topics (see Table 2) were mapped to the existing constructs and factors (see Table 1). Figures 7a and 7b show the mapping with the 5C model and HBM, respectively. In Figures 7a and 7b, the left-most component represents Model 5C or HBM. Moving toward the right, the second component corresponds to the constructs, and the third component represents its factors. Finally, the rightmost component is the topic. The size of each topic represents percentages of user discussions on the Twitter platform among the whole dataset. Percentages of user discussion per topic are obtained from Table 2. The bigger the size, the more discussions on the topic and reflects users' interest in the topic. Each color signifies the topic mapped to the same factor.

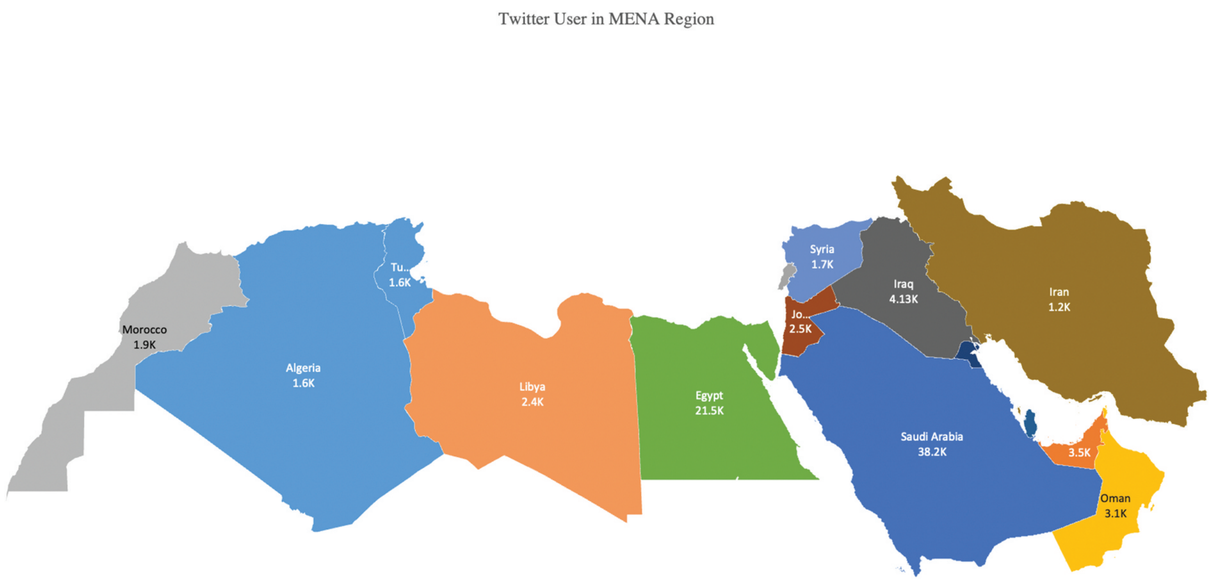

Figure 6. Twitter users in MENA region.

Table 2. Topics extracted from Arabic tweets.

Topics	Keywords Translated from Arabic	PT* (%)	One Example Tweet Translated from Arabic to English
1. Trust in vaccine effectiveness (+)	effective, successful, efficient, positive effect	3.2	My family and I have been vaccinated. I am keen that everyone takes the vaccine because I have confidence in the effectiveness of the vaccine
2. Trust in vaccine safety (-)	safe, reliable, harmless	1.1	I took the Covid vaccine in the form of two doses of the Pfizer type and I have complete confidence in the safety of the COVID-19 vaccine and encourage the community's most vulnerable and eligible members to get the vaccine.
3. Trust in the role of healthcare professionals (+)	doctor, nurse, medical staff	1.9	According to the doctor's suggestion, I took the first dose of the BioNTech vaccine with Pfizer in Germany.
4. Role of policymaker in ensuring vaccines (+)	policymaker, politics, health ministry, WHO	4.3	Kuwait is one of the first countries to receive the vaccine, whatever the cost, knowing that our government has spared its citizens all the best
5. Trust in the role of vaccine manufacturer (+)	pfizer, johnson & johnson, Moderna, astrenzenca, sinopharm, sputnik-v	3.3	It is worth noting that the COVID-19 vaccine produced by Moderna is already being used in many countries around the world, as it has been approved as a safe and effective vaccine by the US, the European Union and the UK
6. Risk perception of COVID-19 (±)	risk of infection, risk of coronavirus, infection	1.3	The Corona virus is like many viruses that we coexisted with, and we did not take a vaccine for it. It is possible to recover at a high rate to the presence of some proven effective drugs and approved treatment. But a new vaccine technology has a lot of confusion and it is impossible to risk our health
7. Past vaccine experiences (±)	Influenza, seasonal flu, malaria, HCV vaccine, polio	1.5	I had taken seasonal vaccines last year. No risk in vaccine. Vaccination is a protection for the elderly and its effectiveness has been proven just like the polio vaccine and other seasonal vaccines (prevention is better than cure)
8. Herd immunity (+)	herd, immunity, immune	1.0	Virologist: Herd immunity against the Delta_strain is formed after vaccinating 80% of the population. Failure to reach this percentage helps the emergence of new mutants that overcome vaccines
9. Vaccine schedule (±)	schedule, appointment, registration, booking	1.5	I am trying to book an appointment to vaccinate children, and doctor says that there are no appointments available for me
10. Willingness to pay for health insurance (±)	health insurance, insurance cost, monthly salary	1.8	Covid vaccines are completely free for everyone in the Kingdom, citizen and resident. Everyone does not need to pay anything or obtain health insurance to get vaccinated.
11. Influences by information sources (+)	print media, twitter, facebook, newspaper	1.7	We advise everyone to take the initiative to take the vaccination, abide by the necessary preventive measures, not be drawn into rumors, and ensure access to correct information from reliable and approved sources.
12. Willingness to accept or refuse vaccine (±)	accept, uptake, refuse, reject, vaccinating	1.4	I am not afraid of vaccination, and I have my daughters who are young, so I don't want them to be vaccinated or examined, and so I refused to for them
13. Precautionary measures (+)	precautionary, mask, preventive, washing hands, social distance	2.1	Everyone must adhere to the instructions for social distancing and preventive measures because the possibilities of infection are still present

(*Continued*)

Table 2. (Continued).

Topics	Keywords Translated from Arabic	PT* (%)	One Example Tweet Translated from Arabic to English
14. Vaccine perception in critical health conditions (-)	hearth inflammation, diabetes, AIDS, cancer	1.1	Flu vaccination is highly recommended for patients with diabetes, chronic bronchitis, and heart patients due to high mortality rates in these groups
15. Vaccine reduces complications (+)	fever, pain, coughing, complication	2.6	The vaccine protects the people around you who are most at risk of contracting the disease and suffering from its complications
16. Vaccine may cause to death (-)	die, death	1.2	The first death caused by blood clotting due to the AstraZeneca vaccine. I think the cases are actually more than that.
17. Vaccine side effects (-)	side effects, blood clot, problems of vaccine, symptoms, long-term, short-term	3.2	All vaccines and medicines have side effects, including the currently approved COVID-19 vaccines.
18. Rapid development of vaccine (-)	rapid, fast development	0.7	Follow most microbiologists, they are not against vaccines. But they are against a rapid vaccine developed in just one year. Any vaccine developed over the years to ensure its safety.
19. Vaccines is new (-)	mRNA, new technology, new vaccine	0.6	The vaccination is still new and needs enough research to prove that it does not have any harmful effects. Especially since vaccinating children against Covid did not take its time like vaccinating adults.
20. Efficacy of vaccines against new variants (+)	variants, mutation, omicron, delta	1.9	Good news about the effectiveness of the Pfizer vaccine against the new genetic mutations that occurred in the Corona virus and made it more capable of spreading
21. Vaccine is required to resume normal life (+)	closure of schools, education ministry, online classes, exam, market	4.2	Obtaining a vaccine is only a first step on the way to returning to normal life, as the vaccine must be highly effective in a large number of the population, to make sure that the pandemic is actually receding
22. Vaccines for pregnant woman (-)	pregnant, breastfeeding, fertility, childcare	0.6	We do not advise pregnant and lactating women to take the vaccine; Vaccinations are recommended to be given to women before pregnancy
23. Vaccines for pregnant woman (-)	PCR, vaccine certificate, coronavirus test, umrah, hajj	4.7	Two doses of the Pfizer vaccine have been taken. Do I need to do a PCR test to travel to Egypt?
24. Influence by conspiracy theories (-)	conspiracy theories, pig, forbidden, electronic chip,	2.0	You will soon be injected with a vaccine containing an 'electronic chip' the size of a grain of rice
25. Influence by religious person (+)	muslim, christian, pope, hindu	1.2	The International Union of Muslim Scholars issues a fatwā on the permissibility of the Coronavirus vaccine

(±, -) topic related to vaccine hesitancy/barriers, (+) topic related to positive attitude/intention to vaccination. PT* - Percentages of Topic.

Mapped Twitter topics to the 5C model

The 5C model comprises five constructs: Confidence, Complacency, Constraints, Calculation, and Collective responsibility. 25 topics have been mapped to 15 factors shown in Figure 7a.

Confidence. Eight topics extracted from Twitter data were mapped to the five factors of the confidence construct. The topic "willingness to accept or refuse vaccines" ($n = 1.4\%$) can be mapped to the factor "attitudes toward vaccines." Twitter users discussed their opinions regarding vaccine acceptance or hesitance/refusal. Tweets related to vaccine acceptance are reported as positive (+) and vaccine hesitance are marked as negative (-). The topic percentages for each factor report the user's engagement for specific concerns.

Three topics, namely "trust in vaccines effectiveness" ($n = 3.2\%$), "vaccines reduce complications" ($n = 2.6\%$), and "efficacy of vaccines against new variants" ($n = 1.9\%$), represent a positive belief in vaccines. These topics can be mapped to the factor "belief about medicine: benefits (+)." Twitter users expressed confidence in the effectiveness of vaccines, emphasizing their role in producing immunity, preventing infection, and aiding recovery. Vaccines were also seen as protective measures for those at risk and capable of reducing complications. Furthermore, users highlighted that vaccines remain effective against new variants of the Coronavirus, including delta, delta plus, and omicron.

The topic "trust in vaccine safety" (1.1%) can be mapped to the factor "belief about medicine: harms (-)." Users were concerned about the safety of vaccines. Some users discussed that the vaccine passed the safety concerns, mentioning their successful completion of clinical trials and encouraging the Twitter community to get vaccinated.

Two topics, "trust in the role of vaccine manufacturers" ($n = 9.7\%$) and "trust in the role of healthcare professionals" ($n = 1.9\%$), can be mapped to the factor "trust in the role of public authorities (+)" Healthcare professionals, such as doctors and nurses, were regarded as frontline defenders against COVID-19, and users followed their vaccination guidelines. Additionally, users believed that the vaccine manufacturers (e.g., Pfizer, Moderna, AstraZeneca, Sinopharm, and Sputnik-V) ensured the safety and effectiveness of vaccines before delivering them to communities.

The topic "influence by conspiracy theories ($n = 2.0\%$)" can be mapped to the factor "conspiracy mentality (-)." Several words circulated indicating the existence of conspiracy theories on Twitter, like hidden parts/electronic chips injected through the vaccine. The most significant of these concerns was Bill Gates, the Microsoft founder, developing the vaccine to implant robots inside the body to remove a third of the world's population. This factor negatively correlates with vaccination confidence because it degrades user beliefs. Overall, these findings from Twitter data reflect user discussions on the trust in vaccines and the role of different stakeholders.

Complacency. Six topics have been mapped to three factors of complacency. The Twitter topic "risk perception of COVID-19 ($n = 1.34\%$)" can be mapped to the factor "perceived threat due

Figure 7. (a) Mapping of 5C model with Twitter topics. From the left side: level 1 is the 5C model. Level 2 is the five constructs in the 5C model. Level 3 shows the factors. The rightmost (level 4) are the topics extracted from Twitter data. (b) Mapping of HBM with Twitter topics. From the left side: level 1 is the HBM. Level 2 is the five constructs in the HBM. Level 3 shows the factors. The rightmost (level 4) are the topics extracted from Twitter data.

to infectious disease (+) ." Many Twitter users in the study expressed concerns about their risk of contracting COVID-19 if they were not vaccinated early. They emphasized the importance of providing vaccines as a crucial measure to prevent infection. This attitude is positively mapped to the factor because it promotes vaccination.

The topic "vaccine may cause death ($n = 1.2\%$)" can be mapped to the factor "consideration of future consequences (-)." Some Twitter users were worried that vaccines might cause blood clotting and increase the chance of respiratory diseases, consequently causing death. This attitude is negatively proportioned to the vaccine acceptance behavior.

Three topics, namely "vaccine perception in critical health conditions ($n = 1.1\%$)," "vaccines for a pregnant woman ($n = 0.6\%$)," and "vaccine side effects ($n = 3.2\%$)" can be mapped to the factor of the perceived risk associated with vaccines. Twitter users in the study were worried about the potential risk after being vaccinated. Particularly, individuals with serious health conditions such as diseases such as AIDS, cancer, inflammatory diseases, and diabetes were identified as the most vulnerable to COVID-19 infection. Twitter users in the study stressed the importance of ensuring vaccines for people in critical health conditions. This topic is positively correlated to vaccine attitudes. Many Twitter users discussed that the COVID-19 vaccine might harm pregnant women and tweeted to avoid vaccinating pregnant women. While Twitter pioneers circulated a rumor that the vaccine causes sterility in women, expectant mothers were advised to be vaccinated after they gave birth. Thus, vaccinating pregnant women is negatively correlated with vaccine attitudes, and Twitter users in the study raised this concern. The short-term side effects of vaccines were mentioned as muscle pain, fatigue, fever, and sneezing; however, the potential long-term of vaccines were unknown. Many users were afraid of getting vaccinated due to side effects, and so it is negatively correlated to vaccine attitudes.

Constraints. Three topics can be mapped to the three factors in the constraints construct. Topic "willingness to pay for health insurance ($n = 1.8\%$)" can be mapped to the factor "affordability and willingness to pay (+)." Twitter users in the study frequently asked whether health insurance is required for vaccination and how much they have to pay. In some countries, having health insurance was required for immunization. However, many people did not have health

Figure 7. (Continued).

insurance and were not willing to pay for it unless their company paid for it. When the vaccine was made accessible to everyone by the government, Twitter users in the study posted about the price the government had to pay to buy vaccine doses. This factor is positively hypothesized with vaccination.

Topic "role of policymaker in ensuring vaccines ($n = 4.3\%$)" can be mapped to the factor "geographical accessibility (+)." Geographical accessibility refers to the availability and ease of access to vaccines in various demographic regions of a country through designated vaccination centers. The health ministry of different countries plays a vital role in arranging a sufficient number of vaccines for the nationals and residents. The policy-makers have initiated social campaigns to promote vaccine literacy among the population, emphasizing the importance of vaccination.

Another topic, "vaccine schedule ($n = 1.5\%$)" can be mapped to the factor "perceived time pressure." During the COVID-19 pandemic, everyone had to book an appointment through applications or websites to get a schedule for the COVID-19 vaccine. There are different applications (e.g., Tawakkalna for Saudi Arabia) for vaccine registration and vaccine status checking in different countries. Getting a schedule for the COVID-19 vaccine was difficult due to the substantial number of people who needed to be vaccinated. This topic is negatively correlated with the factor "vaccine attitudes."

Calculation. The engagement of individuals in information searching calculates vaccine decisions by evaluating the risk of infection and vaccine benefits. The calculation construct refers to two factors: "influential factors in vaccine decision" and "risk calculation of vaccines ($n = 1.2\%$)."

Two topics, "influences by information sources ($n = 1.7\%$)" and "influences by religious person ($n = 1.2\%$)" can be mapped to the factor "influential factors in vaccine decision (+)" because these two topics populate vaccine decisions. An authentic source of information gains the trust level of the public quickly. Throughout the COVID-19 pandemic, the health ministry of different countries worldwide provided continuous updates on infection rates, deaths, recoveries, and vaccine coverage. Twitter users in the study followed these verified news sources to monitor coronavirus news, which may positively impact their beliefs. However, circulations of misinformation and rumors about vaccines in social media spread fast and degraded public beliefs. Similar findings were also observed that people searching for extensive information on social media could have biased vaccine decisions.[78]

Influential leaders, such as religious figures, hold significant sway over public attitudes toward vaccination. Muslim scholars, for instance, issued statements indicating no conflicts with vaccine use, while leaders from other religions, like the Pope, encouraged their followers to accept vaccines and support one another during the pandemic. The impact of these influential leaders on public attitudes toward vaccination cannot be overstated.

Three topics, "past vaccine experience ($n = 1.5\%$)," "vaccine is new ($n = 0.7\%$)" and rapid development of vaccines ($n = 0.6\%$) can be mapped to the factor "risk calculation of vaccine." Twitter users in the study consider various factors (e.g., new vaccines and vaccine development) while making decisions about vaccination for themselves and their children. The COVID-19 vaccine was developed with new technology, such as mRNA technology, and many Twitter users in the study complained that the development procedure was faster compared to other vaccines and didn't pass through rigorous testing. This attitude was negatively correlated with vaccine decisions. Also, individual or family member experiences with past vaccines could be positive or negative, which may reflect in the next vaccination decision.

Collective responsibility. It refers to the willingness to protect others by obtaining herd immunity in the community by accepting vaccines. This construct discusses two factors: collectivism and communal orientation. Collectivism discusses the involvement of everyone in solving the problem, whereas, communal orientation discusses the responsibility of everyone.[14]

Two topics, "vaccine is required to resume normal life ($n = 4.2\%$)" and "vaccine is required for travel and access to work ($n = 4.7\%$)" can be mapped to the factor "collectivism." During the COVID-19 pandemic, everyday life was significantly disrupted, with people having to work from home and schools, colleges, and universities being closed. Travel restrictions were imposed, and individuals were required to provide vaccine certificates and negative COVID-19 reports for travel purposes. On Twitter, users engaged in discussions about the importance of widespread vaccination as a means to restore normalcy in daily life. These attitudes are positively associated with vaccine decisions.

Two topics, "herd immunity ($n = 1\%$)" and "precautionary measures ($n = 2.1\%$)" can be mapped to the factor "communal orientation" because everyone's responsibility in society is associated positively with these two topics. Twitter users encouraged others to obtain immunity by vaccinating because it can protect people at risk of contracting the disease and its complications. Also, they highlighted that vaccination is a catalyst to increase immunity to viruses and a safe way to prevent infection. The precautionary measures (e.g., washing hands, taking a preventive vaccination, wearing a mask, avoiding shaking hands, and covering mouths while coughing) were the most important ways to avoid the spread of coronavirus. People were also advised to maintain a social distance and avoid crowding. Twitter users expected to return to everyday life by sharing valuable information, so people became aware of the coronavirus. The MoH of different countries launched various social awareness campaigns to encourage people to protect the community from mass infection.

Mapping Twitter topics to the HBM

The HBM consists of five constructs: perceived benefits, perceived susceptibility, perceived severity, perceived barriers, and cues to action. In this study, 19 topics have been mapped to the five constructs and 15 factors HBM (see 7(b)).

Perceived benefits. Increased perceived benefits of the COVID-19 vaccine considerably decrease vaccine hesitancy. This construct comprises three factors: perceived vaccine effectiveness, decrease in infection rate, and reduction of complications.

The topic "trust in vaccine effectiveness benefits ($n = 3.2\%$)" can be mapped to the factor "perceived vaccine effectiveness" because Twitter users discussed how vaccines are effective against infectious diseases. They posted how vaccines prevent coronavirus and hinder the spread of the virus. Furthermore, the topic "efficacy of vaccines against new variants ($n = 1.9\%$)" can be mapped to the factor of a "decrease in infection rate." The mutation of new variants (e.g., delta, omicron) caused an increased number of infection cases and deaths and spread rapidly throughout the world. However, users highlighted that receiving the second dose and booster dose of the vaccine proved effective in combating these new variants.

Twitter users emphasized that vaccines help reduce infection rates by providing immunity. They shared posts stating that people who got the vaccine are less at risk of contracting a coronavirus. The topic "vaccine reduces complications of coronavirus ($n = 2.6\%$)" can be mapped to the "reduction of complications." Twitter users posted that vaccines reduce complications such as illness, breathing problems, coughing, and fever if infected by the coronavirus. They suggested getting COVID-19 vaccines as they contribute to decreasing the severity of diseases and minimizing hospitalizations.

Perceived susceptibility. Perceived susceptibility encompasses the risk perception of infection and the risk perception of vaccines. It consists of two factors: perceived risk of infection and perceived risk of vaccines, which are negatively correlated to vaccination.

Two topics, "risk perception of COVID-19 ($n = 1.3\%$)" and "vaccine perception in critical health conditions ($n = 1.1\%$)," can be mapped to the factor of "perceived risk of infection." Twitter users expressed concerns about the risk of contracting COVID-19 as the virus rapidly spread worldwide. They also highlighted that individuals with preexisting conditions such as heart inflammation, cancer, AIDS, and other diseases were particularly vulnerable to coronavirus infection. Twitter users demanded early access to vaccines to protect themselves.

Furthermore, two topics, "trust in vaccine safety ($n = 1.1\%$)" and "vaccines for pregnant women ($n = 0.6\%$)," discuss the potential risks associated with vaccines and can be mapped to the factor of "perceived risk of vaccines." Twitter users expressed concerns about vaccine safety and the lack of sufficient evidence regarding the safety of COVID-19 vaccines for pregnant and breastfeeding women. As a precautionary measure, it was suggested to avoid vaccinating pregnant women. Although rumors circulated on Twitter suggesting that vaccines could cause sterility in women, expectant mothers were advised to consider vaccination after giving birth.

Perceived severity. This construct pertains to individuals' concerns about severe negative consequences. It comprises two primary factors: serious complications of vaccines and social and financial consequences.

The topic "vaccine may cause death ($n = 1.2\%$)" can be mapped to the factor of "serious complications of vaccines." On Twitter, users discussed cases where vaccines were associated with deaths. There were rumors circulating on the platform suggesting that COVID-19 vaccines led to blood clotting, heart attacks, and, ultimately, deaths. However, it is important to note that none of this information has been verified, and it has spread widely on Twitter. Such misinformation negatively impacts vaccine decisions.

The COVID-19 pandemic significantly disrupted both individual and social aspects of life. Two topics, "vaccine is required to resume normal life ($n = 4.7\%$)" and "vaccine is required for travel and access to work," can be mapped to the factor of "social and financial consequences." Twitter users in the study expressed their desire to overcome the losses and challenges caused by the pandemic and return to normalcy in their personal and professional lives. This attitude is positively related to vaccination.

Perceived barriers. Perceived barriers refer to the obstacles to individuals' feelings about getting vaccines. This construct has four factors: perception of vaccine side effects, experience with vaccination, introduction to new vaccine, and access to vaccine centers. Topic "vaccine side effects ($n = 3.2\%$)" can be mapped to the factor "perception of vaccine side effects" because, in this topic, Twitter users in the study posted about potential short-term and long-term side effects caused by vaccines. The short-term side effects of the COVID-19 vaccines were fatigue, fever, and muscle pain, but the long-term side effects of the COVID-19 vaccines are still unknown. They argued that the vaccines should be tested on animals to ensure no dangerous side effects. The health organization suggested to allergic patients, pregnant women, and children to stay away from vaccines unless the side effects were known. There might be serious or mild side effects after taking vaccines from family members, which negatively impact vaccine decisions (topic past vaccine experience). Another factor in the vaccination decision is that access to vaccination centers was critical due to the crowded environment, and it was difficult to get an appointment for vaccines (vaccine schedule).

The topic "rapid development of vaccines ($n = 0.7\%$) and "vaccine is new ($n = 0.6\%$)" can be mapped to the factor introduction to new vaccines. In this topic, Twitter users in the study criticized the vaccine development process and the vaccine approval process by the WHO. Many Twitter users complained that vaccine development requires a longer time to prove its efficacy and safety.

Cues to action. Cues to action trigger people to change their behavior through recommendation. Two topics can be mapped to three factors: government-recommended vaccines, influential leaders-recommended vaccines, and media-recommended vaccines. All these factors are positively correlated to vaccine decisions.

The topic "role of the policymaker in ensuring vaccines ($n = 4.3\%$)" can be mapped to the factor "government-recommended vaccine." Policymakers, as representatives of the government, make critical decisions regarding vaccine policies in a country. They ensure the availability of vaccines and promote vaccination procedures through campaigns and social awareness programs. The recommendations from influential leaders also play a significant role in shaping public attitudes positively toward vaccinations.

The topic "influence by information sources ($n = 1.7\%$)" is positively associated with vaccine decisions and can be mapped to the factor "media-recommended" vaccines because social media/print media can influence public attitudes toward vaccines. For example, Twitter took the initiative to fight against misinformation about COVID-19 vaccines.[78] However, it is important to note that the topic "influence by conspiracy theories" is negatively associated with vaccine decisions and can reverse the decision of media-recommended vaccines.

User engagements with Twitter topics

The most prominent topics were identified based on interaction rates (see Table 3). The conspiracy theory topic got the highest number of user interactions ($n = 9.4$) and reached more Twitter users because users actively participated in conspiracy theories. Twitter users also interactively discussed the price of vaccines and insurance. Many Twitter users did not have health insurance and asked about the need for insurance for vaccination. Some topics had fewer interaction rates but many likes or retweets. Returning to normal life got the highest number of replies ($n = 2.0$) and likes ($n = 11.0$) because users discussed the closure and opening of schools, colleges, and universities. Role of policymaker in ensuring vaccines gained public likes.

Table 3. Users engagement.

Topics	Mean Likes	Mean Retweets	Mean Reply	Interaction rate (e-09)
Influences by conspiracy theory	9.6	3.9	1.5	9.4
Willingness to pay for health insurance	10.4	4.3	1.3	5.3
Vaccines is new	10.4	4.3	1.4	5.3
Vaccine side effects	8.4	2.5	1.4	4.2
Vaccine is required to resume normal life	7.4	2.4	1.3	3.9
Vaccine Perception in critical health conditions	11.4	4.3	1.7	3.7
Herd immunity	9.9	3.6	1.6	3.5
Trust in vaccine effectiveness	9.1	3.0	1.4	3.5
Vaccines for pregnant woman	10.8	3.9	1.73	3.2
Policymaker role in ensuring vaccine	8.3	2.9	1.4	2.4

Discussion

Principal findings

This study constructed an innovative methodology to extract and analyze vaccine-related discussion on the Twitter platform by utilizing the theoretical framework of the Health Belief Model (HBM) and the 5C model. By analyzing the association of Twitter topics with these theoretical frameworks, the study seeks to understand the availability and diversity of data that can be leveraged to predict behavioral changes. Understanding the specific elements contributing to hesitancy in different contexts allows for developing targeted interventions to address each community's unique concerns and barriers.[79] Similar findings were also observed in study[47] where the authors extracted vaccine uptake factors from Twitter data. However, our study diverged in focus, examining the factors contributing to vaccination while they assessed vaccine acceptance. We categorized each factor either positively or negatively according to their characteristics.[14,75] Additionally, we identified the percentages of each factor engaged in positive or negative discussions about vaccination decision that illustrates Twitter users' interests in a specific topic. The predictors of vaccine intentions encompass a large dimension of factors, and timely analysis can provide close insights into the causes. For instance, a Twitter-based study revealed that people who didn't get the vaccine caused more people to die, resulted in anticipated regret, and tended people to be vaccinated.[41] Analyzing Twitter data requires less time compared to traditional methods. This study can be used as a prototype for future studies that want to use any social media platform for identifying vaccine behavior.

Extraction of vaccine attitudes from Twitter data shows that it can capture all the relevant factors of the 5C model and HBM. Twitter topics possess valuable information on public opinions and beliefs which reflect vaccination decisions. People of different professions like common people, academicians, technologists, politicians, sports, and celebrities use Twitter to share their thoughts without restriction which is an excellent source to evaluate public perceptions.[80] Along with the 5C model, we also looked at other behavioral models (HBM) to map Twitter topics. We did this to observe if Twitter topics can cover all other theoretical frameworks in the COVID-19 vaccine hesitancy model.[25] For example, the Confidence construct in the 5C model showed how user belief related positively toward vaccines through Twitter topics such as trust in vaccine effectiveness, the reduction of complications, and efficacy against new variants. However, concerns about vaccine safety and the influence of conspiracy theories negatively impacted confidence in vaccines decision.

Twitter user demographics reveal a diverse representation in terms of gender and geographic location. Gender is a crucial factor in shaping vaccine behaviors and decision-making. This study reported a higher percentage of male users and a significant portion with an unknown gender. Several factors may contribute to these gender disparities in the MENA region, such as cultural norms, social expectations, access to the internet, digital literacy, and individual preferences.[81] This study relied only on the named entity for gender identification and resulted in a significant number of anonymous gender users, which may be another reason for gender disparities. Anonymous gender may limit analyzing vaccine attitudes for vaccination programs from Twitter data. It is worth mentioning that gender identification in Twitter data is challenging and might not accurately reflect the true distribution. The study encompasses a wide range of users from different countries, especially the MENA region, allowing for a more comprehensive understanding of vaccine decision-making across various cultural contexts. It has been observed that the number of Twitter users in the study varies in different countries according to the population of a country. For example, Saudi has a larger population, and we observed a higher number of Twitter users from this country. Similarly, Qatar has a small population and fewer Twitter users. There might be other factors such as internet access, socioeconomic condition, and cultural diversity. This diversity in gender and geographic distribution provides valuable insights into the global perspectives and experiences related to vaccine decision-making. Apart from the society who don't use Twitter and technology, however, people may use other social media platforms.

Prior to this research, we performed a survey in the Middle East North Africa (MENA) region to identify public attitudes and behavior toward vaccine hesitancy.[28] Our previous survey study served as a preliminary investigation, allowing us to identify key areas of interest, public attitudes, and behaviors toward vaccine hesitancy, which were further explored and analyzed in the current study using Twitter data. We wanted to observe if we were able to get a similar response from Twitter data like the survey. To be transparent with the data, we didn't remove any Tweets in the Arabic language from the analysis. The use of state-of-the-art language models in Bertopic modeling significantly improves its performance and enhances its ability to identify topics.[70,82] Compared to traditional survey methodologies, the data collection and analysis process on the Twitter platform is much faster, allowing us to automate the analysis and complete it in less than one month.

This real-time feedback from users can be highly valuable for policymakers and those involved in developing interventions for vaccination programs. This study highlights the importance of trust in vaccines and vaccine providers. The policymakers and public health professionals should focus on building trust in vaccines and vaccine providers by providing accurate information about vaccines and ensuring that vaccines are safe and effective. It was observed that users who were concerned about vaccine side effects were less likely to be vaccinated. So, the policymaker should address concerns about vaccine side effects by providing accurate information about the risks and benefits of vaccines, and by working to dispel myths and rumors through transparent media. The study also found that people are more likely to be vaccinated if they believe vaccines are necessary to protect themselves and their community from disease. This can be achieved by organizing public awareness program and campaigns. Thus, the study findings can provide valuable insights to the policymaker in curbing vaccine hesitancy.

Theoretical contribution

The findings of this study significantly contribute to understanding two well-known theoretical frameworks, the 5C model and the Health Belief Model (HBM), in the context of vaccine attitudes. By applying these models to analyze vaccine-related content on Twitter, this study provides valuable insights into how these theories can be practically employed to interpret public behavior and perceptions regarding vaccine acceptance and hesitancy. The utilization of the 5C model and HBM enabled a structured analysis of diverse constructs related to vaccine attitudes present in the Twitter data.

5C model

The analysis of Twitter data using the 5C model explores factors contributing to vaccine acceptance and hesitancy.

Confidence and trust. Twitter users were more likely to accept vaccines when they had trust in the safety and effectiveness of the vaccine, as well as confidence in the roles played by healthcare professionals and manufacturers. The positive belief in vaccines was evident, with users emphasizing their role in immunity and prevention. However, the presence of conspiracy theories and misinformation undermines users' trust in vaccine safety and thereby increases vaccine hesitancy.[83] Future research should focus on developing targeted interventions to address misinformation and enhance trust in vaccines and healthcare authorities.

Complacency and risk perception. Twitter users revealed concerns about the perceived threat of infectious disease, future consequences, and perceived risks associated with vaccines. For example, pregnant women were vulnerable to infection due to unforeseen side effects of vaccines, and they were advised to vaccinate after giving birth to the child. This perspective is negatively correlated with vaccine attitudes, indicating hesitancy toward vaccinating pregnant women. Effective communication strategies and targeted interventions are needed to address these concerns and promote vaccination among high-risk groups.[16]

Constraints and accessibility. Twitter users discussed practical barriers to vaccine acceptance and vaccine access. It highlighted the government role in overcoming these barriers by buying vaccines promptly and ensuring widespread accessibility, irrespective of geographical location.[42,84,85] One practical obstacle identified is the requirement for individuals to have health insurance to access vaccines.[20,86] Therefore, addressing issues of affordability and insurance coverage is vital for equitable vaccine distribution. The policymaker's role is not only to make vaccines physically accessible but also to motivate individuals to make informed choices regarding their vaccination. This can positively influence vaccine acceptance and compliance.

Calculation and decision-making. Individuals engaged in information searching to calculate vaccine decisions by evaluating risks and benefits. The role of reliable information sources, influential leaders, and past experiences all played a significant role in shaping public attitudes and behaviors related to vaccination. The spread of misinformation and rumors quickly degraded public trust and created biases in vaccine decisions, therefore emphasizing the need to disseminate accurate news through verified sources.[87]

Collective responsibility and community engagement. Twitter focused on the collective and communal attempt to achieve herd immunity and restore normalcy by widespread vaccination. It emphasized the collective efforts and responsibilities by encouraging vaccination to protect the community. This attitude positively impacted users' beliefs toward vaccines.

HBM model

Utilizing the Health Belief Model, this study discusses Twitter user's concerns on COVID-19 vaccines.

Perceived benefits and susceptibility. Twitter users highlighted the effectiveness of vaccines in preventing infections and reducing complications. Vaccines were found to be effective against new variants and thus controlled the pandemic progression.[88] However, they were concerned about vaccine safety for specific populations, such as pregnant women. The contrasting discussions on perceived benefits and susceptibility affected public opinion in vaccination decisions.

Perceived severity and barriers. Potential severe consequences of vaccine and unverified rumors influenced public opinions and contributed to vaccine hesitancy.[89] Individuals were concerned about vaccine side effects and the rapid development of vaccines. Necessarily targeted interventions should be taken to lighten fears and build trust in vaccine safety and efficacy.

Cues to action. Recommendations from the government, influential leaders, and media played an important role in shaping vaccine decisions. A credible source of information promotes vaccine acceptance. However, the presence of false narratives and conspiracy theories in online media may counteract the positive impact of vaccines. There needs to be an effective intervention to address specific concerns and misconceptions circulated through online and social media.

These theories help to systematically categorize and interpret various dimensions of vaccine-related discussions on a popular social media platform. Moreover, the study extends the utilization of other theoretical frameworks to gain insights into public health-related behaviors on different social media data.

Research implications

The study open the space to perform an in-depth analysis of each factor identified in this study to understand the underlying reasons, beliefs, and emotions behind vaccine hesitancy. For example, this study identified the influence of policymakers and other influential figures in shaping vaccine decisions. Future research can be performed to understand which figures are most influential in different contexts and how their communication strategies impact vaccine decisions. Researchers can evaluate the efficacy of various communication strategies informed by the HBM findings. For instance, strategies that emphasize collective responsibility versus individual benefits or messages detailing the vaccine development process against its real-world results to see which influences people's decision to get vaccinated more. Twitter data has the potential to provide valuable insights into public perspectives toward vaccines. Most of the existing works[50,90-92] focus on the content analysis of tweets, and a few of them[40,47,72] identified the vaccine hesitancy/acceptance factors by mining Twitter data. Some works in Arabic tweets focused on sentiment analysis,[52,60] misinformation in vaccine coverage,[93] and stance analysis on vaccine content.[94] There is a research gap in measuring vaccine hesitancy and vaccine intention by following the behavioral model. This study addresses a research gap by deciphering vaccine attitudes from Twitter data and mapping them to the theoretical model of health behavior. This study shows the potential of mining social media data and developing public health interventions and communication strategies in the context of vaccine acceptance and hesitancy. However, this study focuses only on Arabic Tweets and COVID-19 vaccines, which limits the exploration of other languages' tweets. Further research is needed in different languages and social media platforms to understand vaccine attitudes globally.

Limitations

This study bears several limitations. In this study, we did not classify tweets based on how much each topic agrees or disagrees with the determinant. Although the 5C model and HBM model are well defined, the manual interpretation of the topic may limit mapping in some cases because of the authors' understanding. This study was narrowed down to Arabic Tweets, so generalization of results with other languages may be difficult. For example, users' expressions of sarcasm and slang on social media discussions appear differently depending on the geographical and temporal situation. It's important to acknowledge that there might be some bias in mapping the Twitter topics to the constructs and factors. Despite the authors' agreement on the mapping, there could still be misinterpretations of certain topics, leading to potential biases in the results. Also, we didn't measure the stance (e.g., level of agreement) of attitudes toward vaccines in this study. This study only explores the availability of content related to vaccine attitudes. Even though there is a limitation to administering the scale as a survey in terms of Arabic tweets, this study opens a way for the researchers to find a complementary option to survey for analyzing attitudes toward vaccines. This study provides a reasonable basis for analyzing social media data.

Conclusion

This research represents an alternative approach to understanding public attitudes and behaviors related to vaccine acceptance and hesitancy by harnessing the power of social media data, specifically Twitter. This study successfully categorized and analyzed a diverse range of factors contributing to vaccine hesitancy by utilizing established theoretical frameworks such as the 5C model and the Health Belief Model (HBM). The results of this study emphasize the significant importance of trust, confidence, and proficient communication in encouraging the acceptance of vaccines. Examining discussions within these models gives us valuable insights into the underlying reasons, beliefs, and emotions behind vaccine-related decisions. Further research is needed to delve into the factors identified in this research to develop interventions for vaccination programs.

Acknowledgments

Open Access funding provided by Qatar National Library.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Md. Rafiul Biswas http://orcid.org/0000-0002-5145-1990
Zubair Shah http://orcid.org/0000-0001-7389-3274

Author contributions

Conceptualization: Z.S. Research and Writing: M.R.B. and Z.S. Analyses and Figures: M.R.B. Topic modeling and labeling: M.R.B. Twitter Data Acquisition: M.R.B. Discussion: M.R.B. and Z.S. Manuscript review: M.R.B. and Z.S.

Data availability statement

Vaccine-related Twitter data is available on request for research purposes only and can be requested by emailing author M.R.B. rafiulbiswas@gmail.com

References

1. Gudding R. Vaccination as a preventive measure. Fish Vaccination. 2014;9780470674:12–21. doi:10.1002/9781118806913.ch2.
2. Cai C, Peng Y, Shen E, Huang Q, Chen Y, Liu P, Guo C, Feng Z, Gao L, Zhang X, et al. A comprehensive analysis of the efficacy and safety of COVID-19 vaccines. Mol Ther. 2021;29(9):2794–805. doi:10.1016/j.ymthe.2021.08.001.
3. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav. 2021;5 (7):960. doi:10.1038/s41562-021-01172-y.
4. Hernandez RG, Hagen L, Walker K, O'Leary H, Lengacher C. The COVID-19 vaccine social media infodemic: healthcare providers' missed dose in addressing misinformation and vaccine hesitancy. Hum Vaccines Immunother. 2021;17(9):2962–4. doi:10.1080/21645515.2021.1912551.
5. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and

globalized infectious diseases. Hum Vaccines Immunother. 2020;16(11):2586–93. doi:10.1080/21645515.2020.1780846.
6. Burki T. Outbreak of coronavirus disease 2019. Lancet Infect Dis. 2020;20(3):292–3. doi:10.1016/S1473-3099(20)30076-1.
7. Williams WW. Immunization policies and vaccine coverage among adult. The risk for missed opportunities. Ann Intern Med. 1988;108(4):616–25. doi:10.7326/0003-4819-108-4-616.
8. World Health Organization. Ten threats to global health in 2019. Geneva Switzerland: World Health Organisation (WHO); Published 2019 [accessed Sep 20 2022.]. https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019.
9. Bratic JS, Seyferth ER, Bocchini JA. Update on barriers to human papillomavirus vaccination and effective strategies to promote vaccine acceptance. Curr Opin Pediatr. 2016;28(3):407–12. doi:10.1097/MOP.0000000000000353.
10. Downs LS, Scarinci I, Einstein MH, Collins Y, Flowers L. Overcoming the barriers to HPV vaccination in high-risk populations in the US. Gynecol Oncol. 2010;117(3):486–90. doi:10.1016/j.ygyno.2010.02.011.
11. Gonik B. Strategies for fostering HPV vaccine acceptance. Infect Dis Obstet Gynecol. 2006;2006:1–4. doi:10.1155/IDOG/2006/36797.
12. Janz NK, Becker MH. The health belief model: a decade later. Heal Educ Behav. 1984;11(1):1–47. doi:10.1177/109019818401100101.
13. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process. 1991;50(2):179–211. doi:10.1016/0749-5978(91)90020-T.
14. Betsch C, Schmid P, Heinemeier D, Korn L, Holtmann C, Böhm R. Beyond confidence: development of a measure assessing the 5C psychological antecedents of vaccination. PloS One 2018;13. doi:10.1371/journal.pone.0208601.
15. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Heal. 1998;13(4):623–49. doi:10.1080/08870449808407422.
16. Rancher C, Moreland AD, Smith DW, Cornelison V, Schmidt MG, Boyle J, Dayton J, Kilpatrick DG. Using the 5C model to understand COVID-19 vaccine hesitancy across a national and South carolina sample. J Psychiatr Res. 2023;160(January):180–6. doi:10.1016/j.jpsychires.2023.02.018.
17. Wismans A, Thurik R, Baptista R, Dejardin M, Janssen F, Franken I. Psychological characteristics and the mediating role of the 5C model in explaining students' COVID-19 vaccination intention. PloS One. 2021;16(8 August):1–23. doi:10.1371/journal.pone.0255382.
18. Phillips R, Gillespie D, Hallingberg B, Evans J, Taiyari K, Torrens-Burton A, Cannings-John R, Williams D, Sheils E, Ashfield-Watt P, et al. Perceived threat of COVID-19, attitudes towards vaccination, and vaccine hesitancy: a prospective longitudinal study in the UK. Br J Health Psychol. 2022;27(4):1354–81. doi:10.1111/bjhp.12606.
19. Wong MCS, Wong ELY, Huang J, Cheung AW, Law K, Chong MK, Ng RW, Lai CK, Boon SS, Lau JT, et al. Acceptance of the COVID-19 vaccine based on the health belief model: a population-based survey in Hong Kong. Vaccine. 2021;39(7):1148–56. doi:10.1016/j.vaccine.2020.12.083.
20. Lin Y, Hu Z, Zhao Q, Alias H, Danaee M, Wong LP. Understanding COVID-19 vaccine demand and hesitancy: a nationwide online survey in China. PLoS Negl Trop Dis. 2020;14(12):e0008961. doi:10.1371/journal.pntd.0008961.
21. Santos AJ, Kislaya I, Machado A, Nunes B. Beliefs and attitudes towards the influenza vaccine in high-risk individuals. Epidemiol Infect. 2017;145(9):1786–96. doi:10.1017/S0950268817000814.
22. Teitler-Regev S, Shahrabani S, Benzion U. Factors affecting intention among students to be vaccinated against A/H1N1 influenza: a health belief model approach. Adv Prev Med. 2011;2011:1–8. doi:10.4061/2011/353207.
23. Guidry JPD, Laestadius LI, Vraga EK, Miller CA, Perrin PB, Burton CW, Ryan M, Fuemmeler BF, Carlyle KE. Willingness to get the COVID-19 vaccine with and without emergency use authorization. Am J Infect Control. 2021;49(2):137–42. doi:10.1016/j.ajic.2020.11.018.
24. Patwary MM, Bardhan M, Disha AS, Hasan M, Haque MZ, Sultana R, Hossain MR, Browning MH, Alam MA, Sallam M. Determinants of COVID-19 vaccine acceptance among the adult population of Bangladesh using the health belief model and the theory of planned behavior model. Vaccines. 2021;9(12):1–19. doi:10.3390/vaccines9121393.
25. Hossain MB, Alam MZ, Islam MS, Sultan S, Faysal MM, Rima S, Hossain MA, Mamun AA. Health belief model, theory of planned behavior, or psychological antecedents: what predicts COVID-19 vaccine hesitancy better among the Bangladeshi adults? Front Public Health. 2021;9(August):1–0. doi:10.3389/fpubh.2021.711066.
26. MacDonald NE. Vaccine hesitancy: definition, scope and determinants. Vaccine. 2015;33(34):4161–4. doi:10.1016/j.vaccine.2015.04.036.
27. Hidayana I, Amir S, Pelupessy DC, Rahvenia Z, Probandari A. Using a health belief model to assess COVID-19 vaccine intention and hesitancy in Jakarta, Indonesia. PLOS Glob Public Heal. 2022;2(10):e0000934. doi:10.1371/journal.pgph.0000934.
28. Biswas MR, Ali H, Ali R, Shah Z. Influences of social media usage on public attitudes and behavior toward COVID-19 vaccine in the Arab world. Hum Vaccin Immunother. 2022 Jun 7;2074205. doi:10.1080/21645515.2022.2074205.
29. Zhang M, Qi X, Chen Z, Liu J. Social bots' involvement in the COVID-19 vaccine discussions on Twitter. Int J Environ Res Public Health. 2022;19(3). doi:10.3390/ijerph19031651.
30. Chen J, Li K, Zhang Z, Li K, Yu PS. A survey on applications of artificial intelligence in fighting against COVID-19. ACM Comput Surv. 2022;54(8): doi:10.1145/3465398.
31. Larson HJ, Jarrett C, Schulz WS, Chaudhuri M, Zhou Y, Dube E, Schuster M, MacDonald NE, Wilson R. Measuring vaccine hesitancy: the development of a survey tool. Vaccine. 2015;33(34):4165–75. doi:10.1016/j.vaccine.2015.04.037.
32. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Heal. 2018;6(June):1–18. doi:10.3389/fpubh.2018.00149.
33. Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol. 1990;43(1):87–91. doi:10.1016/0895-4356(90)90060-3.
34. Adair JG. The Hawthorne effect: a reconsideration of the methodological artifact. J Appl Psychol. 1984;69(2):334–45. doi:10.1037/0021-9010.69.2.334.
35. Biswas MR, Alzubaidi MS, Shah U, Abd-Alrazaq AA, Shah Z. A scoping review to find out worldwide COVID-19 vaccine hesitancy and its underlying determinants. Vaccines. 2021;9(11). doi:10.3390/vaccines9111243.
36. Sallam M. Covid-19 vaccine hesitancy worldwide: a concise systematic review of vaccine acceptance rates. Vaccines. 2021;9(2):1–15. doi:10.3390/vaccines9020160.
37. Aw J, Seng JJB, Seah SSY, Low LL. COVID-19 vaccine hesitancy—A scoping review of literature in high-income countries. Vaccines. 2021;9(8):900. doi:10.3390/vaccines9080900.
38. Dubé E, Gagnon D, MacDonald N, Bocquier A, Peretti-Watel P, Verger P. Underlying factors impacting vaccine hesitancy in high income countries: a review of qualitative studies. Expert Rev Vaccines. 2018;17(11):989–1004. doi:10.1080/14760584.2018.1541406.
39. Signorini A, Segre AM, Polgreen PM. 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(5). doi:10.1371/journal.pone.0019467.
40. Hou Z, Tong Y, Du F, Lu L, Zhao S, Yu K, Piatek SJ, Larson HJ, Lin L. Assessing COVID-19 vaccine hesitancy, confidence, and public engagement: a global social listening study. J Med Internet Res. 2021;23(6):1–11. doi:10.2196/27632.
41. Pokharel M, Lillie HM, Nagatsuka K, Barbour JB, Ratcliff CL, Jensen JD. Social media narratives can influence vaccine intentions: the impact of depicting regret and character death. Comput Human Behav. 2023;141. doi:10.1016/j.chb.2022.107612.

42. Catalan-Matamoros D, Santamaria-Ochoa CD, Peñafiel-Saiz C. Message analyses about vaccines in the print press, television and radio: characteristics and gaps in previous research. J Commun Healthc. 2019;12(2):86–101. doi:10.1080/17538068.2019.1614377.
43. Alrubaian M, Al-Qurishi M, Hassan MM, Alamri A. A credibility analysis system for assessing information on twitter. IEEE Trans Dependable Secur Comput. 2018;15(4):661–74. doi:10.1109/TDSC.2016.2602338.
44. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis. 2014;14(2):160–8. doi:10.1016/S1473-3099(13)70244-5.
45. Betsch C, Brewer NT, Brocard P, Davies P, Gaissmaier W, Haase N, Leask J, Renkewitz F, Renner B, Reyna VF, et al. Opportunities and challenges of web 2.0 for vaccination decisions. Vaccine. 2012;30(25):3727–33. doi:10.1016/j.vaccine.2012.02.025.
46. Liu S, Liu J. Understanding behavioral intentions toward COVID-19 vaccines: theory-based content analysis of tweets. J Med Internet Res. 2021;23(5). doi:10.2196/28118.
47. Baj-Rogowska A. Mapping of the COVID-19 vaccine uptake determinants from mining twitter data. IEEE Acces. 2021;9 (i):134929–44. doi:10.1109/ACCESS.2021.3115554.
48. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, Brownstein JS. A case study of the New York city 2012-2013 influenza season with daily geocoded twitter data from temporal and spatiotemporal perspectives. J Med Internet Res. 2014;16(10):e236. doi:10.2196/jmir.3416.
49. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg. 2012;86(1):39–45. doi:10.4269/ajtmh.2012.11-0597.
50. Dyda A, Shah Z, Surian D, Martin P, Coiera E, Dey A, Leask J, Dunn AG. HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among Australian Twitter users. Hum Vaccines Immunother. 2019;15(7–8):1488–95. doi:10.1080/21645515.2019.1596712.
51. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, Jha S, Agrawal M, Dziurzynski LA, Sap M, et al. Psychological language on twitter predicts county-level heart disease mortality. Psychol Sci. 2015;26(2):159–69. doi:10.1177/0956797614557867.
52. Alliheibi FM, Omar A, Al-Horais N. Opinion mining of Saudi responses to COVID-19 vaccines on twitter: a computational linguistic approach. Int J Adv Comput Sci Appl. 2021;12(6):72–8. doi:10.14569/IJACSA.2021.0120610.
53. Opel DJ, Taylor JA, Mangione-Smith R, Solomon C, Zhao C, Catz S, Martin D. Validity and reliability of a survey to identify vaccine-hesitant parents. Vaccine. 2011;29(38):6598–605. doi:10.1016/j.vaccine.2011.06.115.
54. Alhazmi H. Arabic Twitter conversation dataset about the COVID-19 vaccine. Data. 2022;7(11). doi:10.3390/data7110152.
55. Haouari F, Hasanain M, Suwaileh R, Elsayed T. ArCOV-19: the first Arabic COVID-19 Twitter dataset with propagation networks. WANLP 2021 - 6th Arab Nat Lang Process Work Proc Work. 2021;19:82–91.
56. Alqurashi S, Alhindi A, Alanazi E. Large Arabic Twitter dataset on COVID-19. Published online 2020. 2–4. http://arxiv.org/abs/2004.04315.
57. Imran M, Qazi U, Ofli F. TBCOV: two billion multilingual COVID-19 tweets with sentiment, entity, geo, and gender labels. Data. 2022;7(1). doi:10.3390/data7010008.
58. Mubarak H, Hassan S, Chowdhury SA, Alam F. ArCovidVac: analyzing Arabic tweets about COVID-19 vaccination. 2022. http://arxiv.org/abs/2201.06496.
59. Biswas R, Alam T, Househ M, Shah Z. Public sentiment towards vaccination after COVID-19 outbreak in the Arab world. Stud Health Technol Inform. 2022;289:57–60. doi:10.3233/SHTI210858.
60. Alhejaili R, Alhazmi ES, Alsaeedi A, Yafooz WMSSentiment analysis of the COVID-19 vaccine for Arabic tweets using machine learning. 2021 9th Int Conf Reliab Infocom Technol Optim (Trends Futur Dir ICRITO 2021. Published online 2021;7–11.10.1109/ICRITO51393.2021.9596517.
61. TwitterTwitter API for academic research | products | Twitter developer platform. [accessed May 30 2021]. https://developer.twitter.com/en/products/twitter-api/academic-research.
62. Arabic Stop words. [accessed Sep 18 2022]. https://github.com/linuxscout/arabicstopwords.
63. Attia MA Arabic tokenization system. 2007;(June):65. doi:10.3115/1654576.1654588
64. Arabic-names-gender-classification. [accessed Jul 3 2023]. https://github.com/bibs2091/Arabic-names-gender-classification.
65. Sekine S, Ranchhod E, eds. *Named entities: recognition, classification and use*. Vol. 19. John Benjamins Publishing; 2009.
66. Shah Z, Martin P, Coiera E, Mandl KD, Dunn AG. Modeling spatiotemporal factors associated with sentiment on twitter: synthesis and suggestions for improving the identification of localized deviations. J Med Internet Res. 2019;21(5). doi:10.2196/12881.
67. Daud A, Li J, Zhou L, Muhammad F. Knowledge discovery through directed probabilistic topic models: a survey. Front Comput Sci China. 2010;4(2):280–301. doi:10.1007/s11704-009-0062-y.
68. Grootendorst M. Bertopic: neural topic modeling with a class-based TF-IDF procedure. arXiv. 2022;2203.05794(cs.CL). 2022. doi:10.48550/arXiv.2203.05794.
69. Kellert O, Mahmud Uz Zaman M. Using neural topic models to track context shifts of words: a case study of COVID-related terms before and after the lockdown in April 2020. 2022; (April 2020):131–9. doi:10.18653/v1/2022.lchange-1.14.
70. Abuzayed A, Al-Khalifa H. BERT for Arabic topic modeling: an experimental study on BERTopic technique. Procedia CIRP. 2021;189:191–4. doi:10.1016/j.procs.2021.05.096.
71. Ghazy RM, Abd ElHafeez S, Shaaban R, Elbarazi I, Abdou MS, Ramadan A, Kheirallah KA. Determining the cutoff points of the 5C scale for assessment of COVID-19 vaccines psychological antecedents among the Arab population: a multinational study. J Prim Care Community Health. 2021;12. doi:10.1177/21501327211018568.
72. Fues Wahl H, Wikman Erlandson B, Sahlin C, Nyaku M, Benĉina G. Analysis of vaccine messages on social media (Twitter) in scandinavia. Hum Vaccines Immunother. 2022;18 (1). doi:10.1080/21645515.2022.2026711.
73. Kwok KO, Li KK, WEI WI, Tang A, Wong SYS, Lee SS. Influenza vaccine uptake, COVID-19 vaccination intention and vaccine hesitancy among nurses: A survey. Int J Nurs Stud. 2021;114. doi:10.1016/j.ijnurstu.2020.103854.
74. Mercadante AR, Law AVSince January 2020 Elsevier has created a COVID-19 resource centre with free information in English and mandarin on the novel coronavirus COVID- 19. The COVID-19 resource centre is hosted on Elsevier connect, the company ' s public news and information. 2020 Jan.
75. Limbu YB, Gautam RK, Pham L. The health belief model applied to COVID-19 vaccine hesitancy: a systematic review. Vaccines. 2022;10(6). doi:10.3390/vaccines10060973.
76. Coronavirus disease (COVID-19) weekly epidemiological updates and monthly operational updates. [accessed Mar 17 2023]. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
77. Oliver SE, Gargano JW, Scobie H, Wallace M, Hadler SC, Leung J, Blain AE, McClung N, Campos-Outcalt D, Morgan RL, et al. The advisory committee on immunization practices' interim recommendation for use of Janssen COVID-19 vaccine — United States, February 2021. MMWR Surveill Summ. 2021;70(9):329–32. doi:10.15585/mmwr.mm7009e4.
78. Al-Regaiey KA, Alshamry WS, Alqarni RA, Albarrak MK, Alghoraiby RM, Alkadi DY, Alhakeem LR, Bashir S, Iqbal M. Influence of social media on parents' attitudes towards vaccine administration. Hum Vaccines Immunother. 2021;18(1): doi:10.1080/21645515.2021.1872340.

79. Wiysonge CS, Ndwandwe D, Ryan J, Jaca A, Batouré O, Anya BPM, Cooper S. Vaccine hesitancy in the era of COVID-19: could lessons from the past help in divining the future? Hum Vaccines Immunother. 2021;(1):1–3. doi:10.1080/21645515.2021.1893062.
80. Alabdullatif A, Shahzad B, Alwagait E. Classification of Arabic Twitter users: a study based on user behaviour and interests. Mob Inf Syst. 2016;2016. doi:10.1155/2016/8315281.
81. Bastian BL, Metcalfe BD, Zali MR. Gender inequality: entrepreneurship development in the MENA region. Sustain. 2019;11(22). doi:10.3390/su11226472.
82. Grootendorst M Bertopic: neural topic modeling with a class-based TF-IDF procedure. Published online 2022. http://arxiv.org/abs/2203.05794.
83. French J, Deshpande S, Evans W, Obregon R. Key guidelines in developing a pre-emptive COVID-19 vaccination uptake promotion strategy. Int J Environ Res Public Health. 2020;17(16):1–14. doi:10.3390/ijerph17165893.
84. Hsu AL, Johnson T, Phillips L, Nelson TB. Sources of vaccine hesitancy: pregnancy, infertility, minority concerns, and general skepticism. Open Forum Infect Dis. Published online 2021. doi:10.1093/ofid/ofab433.
85. Guntuku SC, Buttenheim AM, Sherman G, Merchant RM. Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States. Vaccine. 2021;39 (30):4034–8. doi:10.1016/j.vaccine.2021.06.014.
86. Harapan H, Wagner AL, Yufika A, Winardi W, Anwar S, Gan AK, Setiawan AM, Rajamoorthy Y, Sofyan H, Vo TQ, et al. Willingness-to-pay for a COVID-19 vaccine and its associated determinants in Indonesia. Hum Vaccines Immunother. 2020;16 (12):3074–80. doi:10.1080/21645515.2020.1819741.
87. Sallam M, Dababseh D, Eid H, Al-Mahzoum K, Al-Haidar A, Taim D, Yaseen A, Ababneh NA, Bakri FG, Mahafzah A. High rates of COVID-19 vaccine hesitancy and its association with conspiracy beliefs: a study in Jordan and Kuwait among other Arab countries. Vaccines. 2021;9(1):1–16. doi:10.3390/vaccines9010042.
88. El-Shabasy RM, Nayel MA, Taher MM, Abdelmonem R, Shoueir KR, Kenawy ER. Three waves changes, new variant strains, and vaccination effect against COVID-19 pandemic. Int J Biol Macromol. 2022;204:161–8. doi:10.1016/j.ijbiomac.2022.01.118.
89. Walach H, Klement RJ, Aukema W. The safety of COVID-19 vaccinations — we should rethink the policy. Vaccines. 2021;9:693.
90. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: a sentiment analysis. Vaccine. 2021;39 (39):5499–505. doi:10.1016/j.vaccine.2021.08.058.
91. Zhang J, Wang Y, Shi M, Wang X. Factors driving the popularity and virality of COVID-19 vaccine discourse on Twitter: text mining and data visualization study. JMIR Public Heal Surveill. 2021;7(12):e32814. doi:10.2196/32814.
92. Lyu JC, Le HE, Luli GK. COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis. J Med Internet Res. 2021;23(6):e24435. doi:10.2196/24435.
93. Al-Hashedi A, Al-Fuhaidi B, Mohsen AM, Ali Y, Gamal Al-Kaf HA, Al-Sorori W, Maqtary N. Ensemble classifiers for Arabic sentiment analysis of social network (Twitter data) towards COVID-19-related conspiracy theories. Appl Comput Intell Soft Comput. 2022;2022:1. doi:10.1155/2022/6614730.
94. Alqurashi T. Stance analysis of distance education in the kingdom of Saudi Arabia during the COVID-19 pandemic using Arabic twitter data. Sensors. 2022;22(3). doi:10.3390/s22031006.