

RESEARCH ARTICLE



Longitudinal analysis of behavioral factors and techniques used to identify vaccine hesitancy among Twitter users: Scoping review

Sulaiman Khan , Md. Rafiul Biswas , and Zubair Shah 

College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

ABSTRACT

While vaccines have played a pivotal role in the fight against infectious diseases, individuals engage in online resources to find vaccine-related support and information. The benefits and consequences of these online peers are unclear and mainly cause a behavioral shift in user sentiment toward vaccination. This scoping review aims to identify the community and individual factors that longitudinally influence public behavior toward vaccination. The secondary aim is to gain insight into techniques and methodologies used to extract these factors from Twitter data. We followed PRISMA-ScR guidelines to search various online repositories. From this search process, a total of 28 most relevant articles out of 705 relevant studies. Three main themes emerged including individual and community factors influencing public attitude toward vaccination, and techniques employed to identify these factors. Anti-vax, Pro-vax, and neutral are the major communities, while misinformation, vaccine campaign, and user demographics are the common individual factors assessed during this reviewing process. Twitter user sentiment (positive, negative, and neutral) and emotions (fear, trust, sadness) were also discussed to identify the intentions to accept or refuse vaccines. SVM, LDA, BERT are the techniques used for topic modeling, while Louvain, NodeXL, and Infomap algorithms are used for community detection. This research is notable for being the first systematic review that emphasizes the dearth of longitudinal studies and the methodological and underlying practical constraints underpinning the lucrative implementation of an explainable and longitudinal behavior analysis system. Moreover, new possible research directions are suggested for the researchers to perform accurate human behavior analysis.

ARTICLE HISTORY

Received 4 September 2023
Revised 16 October 2023
Accepted 29 October 2023

KEYWORDS




Twitter; longitudinal analysis; scoping review; vaccine hesitancy; users' behavior analysis

Introduction

Recently, longitudinal analysis has gained significant attention from healthcare practitioners and psychologists to perform in-depth analyses of users' health and behavioral modifications with time. During this process, the researchers analyze the data collected from the same subjects or participants over a period of time. It is often used to examine changes, trends, or relationships within a specific group or population. Conversely, to the cross-sectional investigations, the longitudinal analysis allows for the investigation of temporal patterns, the exploration of causal relationships, and the examination of individual trajectories. For human behavior analysis, longitudinal evaluations are performed to understand how responses change over time and identify the factors that influence this change. By repeatedly measuring individuals, researchers can capture the within-individual differences in responses.¹ A considerable amount of data spans over time is required to perform longitudinal analysis however, collecting data through surveys, questionnaires, or public stations can indeed be time-consuming and prone to errors. Additionally, when data collection extends over a span of several years, it can become more complex and financially demanding, making it less feasible for researchers.

With the availability of social media platforms, it has now become more accessible for researchers to collect data spans over time and perform longitudinal behavior analysis. The social media tools like Twitter,^a Facebook, Instagram, YouTube, and others facilitate users exchanging ideas, information, knowledge, and other facts among human societies and generate an enormous amount of data. Research shows that regular use of these applications can have both positive and negative effects on human behavior.² With around 400 million registered users sharing their thoughts, Twitter generates a substantial amount of data daily. Twitter users utilize "tweets," which are short messages, to share news, information, and opinions. Additionally, Twitter provides options for users to engage with tweets by liking, commenting, retweeting (reposting a tweet), and sharing them. Kaur et al.,³ synthesized the dynamics and flow of behavioral changes among twitter users during the pandemic.

With several vaccines approved globally, mass vaccination campaigns are currently underway. However, attitudes toward vaccination, specifically vaccine hesitancy, pose a potential threat to achieving sufficient coverage and community immunity. The SAGE Working Group on Vaccine

CONTACT Sulaiman Khan  sukh45452@hbku.edu.qa  College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, P.O. Box: 34110, Education City, Doha, Qatar; Zubair Shah  zshah@hbku.edu.qa

 Supplemental data for this article can be accessed on the publisher's website at <https://doi.org/10.1080/21645515.2023.2278377>.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Hesitancy determined that vaccine hesitancy pertains to the postponement of acceptance or the refusal of vaccination even when vaccination services are readily available.⁴ Factors such as attitudes toward the collective importance, efficacy, side effects, and speed of vaccine development were significant predictors of vaccine hesitancy, while sociodemographic variables such as age, sex, and socioeconomic status only explained a small proportion (10%) of the variance. Additionally, anti-vaccine content is prevalent across social media platforms, posted by a minority of users, but frequently generating greater user engagement than neutral or pro-vaccine content.⁵ Recent studies have also found that social media posts containing vaccine-related misinformation are shared as frequently as those with reliable information.^{6,7} Disinformation campaigns on social media have been associated with drops in vaccination coverage, as measured by annual data on actual vaccination rates from the WHO, and increased levels of negative vaccine discourse on Twitter.⁸

The literature on vaccine hesitancy lacks clear conceptual clarity, as it presents varying definitions of vaccine hesitancy, ranging from a psychological state to different types of vaccination behavior.^{9,10} Additionally, the terminologies 'low uptake,' 'vaccine confidence,' and 'low intention to vaccinate' are often matched with vaccine hesitancy.^{11,12} Several research studies have employed systematic reviews and surveys to synthesize existing literature and pinpoint knowledge gaps. For instance, in the field of sentiment analysis for vaccine hesitancy, Alamoodi and colleagues¹³ conducted a comprehensive analysis. They highlighted the repercussions of vaccine hesitancy across various sectors, including social, medical, public health, and technology science. In a different study, Zhao and his team¹⁴ carried out a review analysis focusing on the impact of COVID-19 on mental health in Australia. Their longitudinal analysis revealed that specific demographic groups, such as young individuals, those with preexisting mental health conditions, and individuals facing financial hardships, experienced more significant declines in their mental well-being. Another valuable contribution came from Bussink-Voorend et al.,¹⁵ who performed a meticulous systematic analysis of literature gathered from PubMed, Embase, and PsycINFO databases. Their work aimed to provide readers with a clear understanding of the concept of vaccine hesitancy. Lastly, Skafle and his team¹⁶ conducted a scoping review of the literature to identify instances of misinformation related to autism and COVID-19 vaccination shared on social media.

Upon reviewing the literature, a common observation is that the majority of existing scoping reviews tend to concentrate on particular areas, such as misinformation, sentiments, mental health, or other healthcare consequences. Additionally, many of these reviews have limitations in terms of geographic scope or their focus on specific vaccines. In contrast, our review aims to synthesize the literature from various research domains, including the longitudinal analysis of user behavior toward vaccination, factors influencing user behavior, and techniques for identifying user attitudes over time using Twitter data. Furthermore, this review aims to evaluate vaccine hesitancy and its underlying factors (individual and

community factors) on Twitter as a social media platform. The findings of this scoping review will provide valuable insights for healthcare administrators and policymakers to understand the factors associated with vaccine hesitancy among different cohorts engaging in Twitter discussions. This understanding will facilitate the planning of vaccination campaigns and help improve the uptake rates of various vaccines. The main theme of this scoping review followed by the research questions are explained below:

- To identify how the Twitter users' behavior changes toward vaccination longitudinally. Also, it aims to enlist the magnified community and individual factors that are influencing user behavior toward vaccines.
- To describe the state-of-the-art techniques used to identify Twitter users' behavior toward vaccination longitudinally. Furthermore, this objective aims to explain different machine learning and software-based methodologies used to perform users' sentiment analysis toward vaccination.
- To find out both the healthcare and social impacts of vaccine hesitancy on population. Moreover, this objective aims to identify how the vaccine hesitancy challenges the healthcare workers and state agencies to fight against outbreaks.

Methodology

PRISMA-ScR guidelines¹⁷ and vaccine hesitancy scoping review framework¹⁸ are followed to accomplish this longitudinal research work. Table 1 of the supplementary file presents the adherence to the PRISMA-ScR guidelines.

Search process

A search was performed on 10 databases: Scopus, PubMed, IEEE Xplore, ACM, Google Scholar, PsycINFO, Ovid, CINAHL, Springer Link, and Cochrane Library, using relevant keywords and search queries. The MEDLINE database was not searched because PubMed provides primary access to references and abstracts to the MEDLINE database. Google Scholar retrieved a large number of studies (30,600) and so the 100 studies are included from the first 10 pages to keep the search results relevant in this review. The search was conducted on only English-language articles. This study focuses on longitudinal vaccine-related content on Twitter data. Twitter debuted on March 21, 2006. The search was limited from 2006 to current. Forward and backward reference list checking was performed to identify the up-to-date studies.

Search terms

The authors of this study developed search queries through discussion and consultation. It aimed to find out all the relevant studies that performed longitudinal analysis on Twitter data toward vaccines. Search terms were chosen based on the intervention ('longitudinal' OR 'retrospective' OR 'prospective') on Twitter platform (*twitter* OR *tweet**). The target outcome was (*vaccin** OR *immuniz** OR *immunis**). An example

Table 1. Studies analysis.

Paper ids	Data sources	Data samples (posts)	Data timeframe	Unique users	Vaccine studied	Task performed	Techniques applied
19	• Twitter	8579728	Feb 2020 – Dec 2020	576894	Covid-19	Analyze Anti-vaxxer's aggressive behavior	<ul style="list-style-type: none"> • RT-Network • LIWC
20	• Twitter	419269	Jan 2020 – Dec 2020	40081	Covid-19	Analyze users' sentiments and emotions	<ul style="list-style-type: none"> • LDA • VADER
21	• Twitter	1197763	Jun 2020 – Jul 2021	N/A	Covid-19	Identify the public opinion on vaccination	<ul style="list-style-type: none"> • Regression • VADER
22	<ul style="list-style-type: none"> • Twitter • Google • Television • Twitter 	637 3000000	31-Dec-2019 to 30-Nov-2020 Jan 2012 to Jun 2015	N/A	Covid-19	Analyze rumors and conspiracy theories	<ul style="list-style-type: none"> • R version 4.0.3 • Factcheck agency
23	• Twitter	3000000	Jan 2012 to Jun 2015	144817	MMR	Motivation behind anti-vaccination movement	<ul style="list-style-type: none"> • NLP • SVM • STATA
24	• Twitter	36554	19-Oct-2020 to 28-Feb-2021	N/A	Covid-19	Identify the racism and potential inequities	<ul style="list-style-type: none"> • LDA • DeepFace
25	• Twitter	4,438,203	5-Mar-2020 to 25-Jan-2021	1153299	Covid-19	Understand public discussion and perception	<ul style="list-style-type: none"> • VADER • R technique
26	• Twitter	436	31-Jan-2021 to 6-Feb-2021	973	Covid-19	Analyze misinformation and hesitancy	<ul style="list-style-type: none"> • LDA • Prism
27	• Twitter	334747	4-Jan-2020 to 14-Dec-2020	3463	Covid-19	Monitor emergency physicians' attitude during pandemic	<ul style="list-style-type: none"> • VADER • R and ANOVA • HLM
28	• Twitter	108307	Jan-2020 to Dec-2020	106412	Covid-19	Analyze rumors and user stance for vaccination	<ul style="list-style-type: none"> • Silhouette width • Louvain method • NLP • Naïve Bayes • Louvain method • LIWC • VADER • NLP • Domain network graph
29	• Twitter	660892	1-Feb-2015 to 5-Mar-2015	269623	MMR	Identify the vocal patterns of anti and pro-vax users	
30	• Twitter	N/A	Nov-2021 to Jan-2022	337	Covid-19	Vaccine misinformation and warning iconography	
31	<ul style="list-style-type: none"> • Twitter • Scientific • Non-sci • Custom scripts 	<ul style="list-style-type: none"> • Sci: 8,612 • Non-sci: 45320 	1-Jan-2016 to 28-Nov-2016 28-Aug-2017 to 9-Oct-2017	6612	MMR	Monitor controversies over MMR vaccine and autism	<ul style="list-style-type: none"> • Louvain method • Tidytext package • Page Rank • BERT • VADER • Pearson values • NLP
32	• Twitter	10710	28-Aug-2017 to 9-Oct-2017	2600	MMR	Analyze social discussion among Dutch users	
33	<ul style="list-style-type: none"> • Facebook • Twitter 	<ul style="list-style-type: none"> • Facebook US: 144,864 UK: 23571 • Twitter US: 98385 UK: 40268 	1-Mar-2020 to 22-Nov-2020	N/A	Covid-19	Monitor vaccine hesitancy among US and UK residents	
34	• Twitter	13093406	Nov-2020 to Mar-2022	N/A	Covid-19	Analyze public vaccine acceptance rate	<ul style="list-style-type: none"> • XLM-R • Spline regression • R version 4.2.1 • IQR • Ensemble model • Kruskal-Wallis • ML • Botomer API v-4 • SPSS version 23 • NodeXL
35	• Twitter	83551	Oct-2013 to Apr-2014	30621	HPV	Analyze public opinion about HPV vaccine	
36	• Twitter	193379	Aug-2014 to Jul-2015	N/A	HPV	Quantify health master's communication	
37	• Twitter	N/A	24-Dec-2020 to 30-Apr-2021	Anti-vax: 4663 Pro-vax: 2583 Bots: 582	Covid-19	Monitoring the role of bots in the dissemination of vaccine information	

(Continued)

Table 1. (Continued).

Paper ids	Data sources	Data samples (posts)	Data timeframe	Unique users	Vaccine studied	Task performed	Techniques applied
³⁸	• Twitter	1.8 M	Jun-2020 to Apr-2021	N/A	Covid-19	analyzing the temporal evolution of emotions	• NLP
³⁹	• Twitter	1834	Nov-2017 to Jan-2018	Parents: 1026	HPV	Assess how social media affects HPV vaccination	• InfoMap
⁴⁰	• Instagram						• SAS version 9.4
	• Facebook						
⁴¹	• Twitter	1431463	1-Jan-2014 to 26-Oct-2018	486116	HPV	Study public behavioral change about HPV vax	• LOESS & HBM
	• Twitter	2868260	Jan-2020 to Apr-2021	N/A	Covid-19	Assess emotional impact on COVID-19 vaccination	• Att-RNN
⁴²	• Twitter	168950	1-Jul-2020 to 6-Jan-2021	142331	Covid-19	Extract misinformation about COVID-19 vaccine	• RoBERTA model
⁴³	• Twitter	956262	Jan-2014 to Dec-2017	N/A	HPV	Characterize health beliefs about HPV vax	• NLP
⁴⁴	• Twitter	2 M	Nov-2020 to Jan-2021	522893	Covid-19	Identify user engagement to anti-vaccination tweets	• MBFC & PCA
⁴⁵	• Twitter	968519	Mar-2021 to Jan-2022	N/A	Covid-19	Sentiment analysis and aspect understanding during Covid-19	• NLP
							• Bi-RNN & HBM
							• RStudio
							• BERT
							• Bi-LSTM
							• SVM
⁴⁶	• Twitter	672133	18-Nov-2020 to 3-Feb-2021	N/A	Covid-19	Monitor user sentiment toward COVID-19	• R language
							• VADER

of search queries for Scopus databases is: *TITLE-ABS KEY ((‘longitudinal’ OR ‘retrospective’ OR ‘prospective’) AND (twitter OR tweet*)) AND (vaccin* OR immuniz* OR immunis*)*. The detailed search terms that are used for different databases are shown in Appendix Search Results excel file.

Inclusion and exclusion criteria

A primary search was performed to find literature about longitudinal vaccine-related topics on Twitter data. Studies considered Twitter data along with other social media data were included in this study. Articles written only in the English language were considered for this study. The included studies are empirical studies, peer-reviewed articles, dissertations, book chapters, and conference proceedings, and studies published from 2006 to the present. In the study setting, there was no limitation on the population’s demographic data, such as age, gender, and nationality.

Articles were excluded that were published before 2006 because Twitter started in 2006. Articles written in languages other than English were not considered. All those longitudinal studies were excluded that analyzed Tweets instead of Twitter users. Studies were excluded that failed to evaluate user(s) sentiments for a certain period of time. Longitudinal analysis on hospitals or other health workers vaccination reports was removed from search results. Also, all those studies were excluded that manually assessed/surveyed users’ sentiments for the vaccine. Systematic reviews, newspapers, magazines, reviews, proposals and posters, non-peer-reviewed articles, only abstracts, and letters to the editor were excluded from the analysis.

Study selection

The study selection procedure comprised four stages: initially, duplicate studies were identified using the automated duplicate detection feature of Rayyan software⁴⁷ and subsequently eliminated. In the following step, two reviewers (SK and MRB) assessed the titles and abstracts of the remaining unique studies and included if they met the study inclusion criteria. Next, the reviewers thoroughly examined the full text of the selected studies. Finally, forward and backward reference lists were checked in the included studies to further observe the relevant studies. Any discrepancies between the reviewers were resolved through discussion with the third author ZS.

Data extraction and synthesizing

To prepare for data extraction from the included studies, an extraction table was created with a column header using an Excel spreadsheet and shared among other authors MRB and ZS. The data extraction sheet was reviewed and updated through discussion. A pilot test was conducted on two studies to ensure data consistency and availability. The extracted data encompassed various aspects, including study characteristics (such as type of paper, authors, authors’ location), attributes of Twitter data (such as tool online source searched (Twitter), online source searched (others), tweet language, tweets language (others) total duration, number of tweets, and number

of users). Vaccine-related topics to data acquisition (such as type of vaccine, topics, emotions, peak time, and concerns). The extracted data are presented and summarized in the tables. The analysis and findings are synthesized using a narrative approach.

Results

This section of the paper outlines the search results and main findings of this scoping review process. It briefly explains the articles accumulation and selection process, key factors influencing user’s behavior toward vaccination, and methodologies used in the literature (included articles), to longitudinally analyze psychological characteristics associated with vaccine hesitancy.

Search results

During our initial articles downloading process, 705 records were obtained that were further screened for removing duplicate entries; we were left with 431 studies for further evaluation. Using our established inclusion and exclusion criteria (see Methods section), we further screened these studies based on their title and abstract and obtained 294 studies for full-text review. After screening these articles based on the contents provided, we obtained only 28 of the most relevant articles for our final evaluation and reviewing process. During this process, we removed the articles that considered only tweets instead of users because we aimed to study users’ behavior toward vaccination, not the tweet contents. Also, we removed those articles that presented other social media sources other than Twitter because our primary concern is to consider only Twitter. It is worth mentioning that we included those articles that considered other social media sources, along with Twitter. Furthermore, we excluded those articles that performed manual surveys, because our primary concern is to analyze the literature that considered Twitter a primary source for longitudinal user’s analysis toward vaccination. The overall search process, articles downloading and screening process, and final pool of most relevant studies selection process are described in Figure 1.

Demographic of the included studies

The overall demographics of the finalized relevant articles are shown in Figure 2. The outer shell represents the references to the included articles, the second last represents the type of vaccines reported in these longitudinal studies, the third last shell represents the publication year, and the inner circle represents the country of the first author (reported in the studies). These vaccines include Measles, Mumps, and Rubella (MMR) vaccine, Human Papillomavirus Vaccine (HPV), and coronavirus disease of 2019 (Covid-19). Most of the studies ($N = 19 \sim 68\%$) reported COVID-19 vaccination, ($N = 5 \sim 18\%$) reported HPV vaccine, and ($N = 4 \sim 14\%$) reported MMR vaccines. Most of the studies ($N = 11 \sim 39\%$) were reported in 2022, while ($N = 9 \sim 32\%$) studies reported during 2021. Among the total 28 studies, 20 studies ($\sim 72\%$) longitudinal studies on user behavior analysis toward

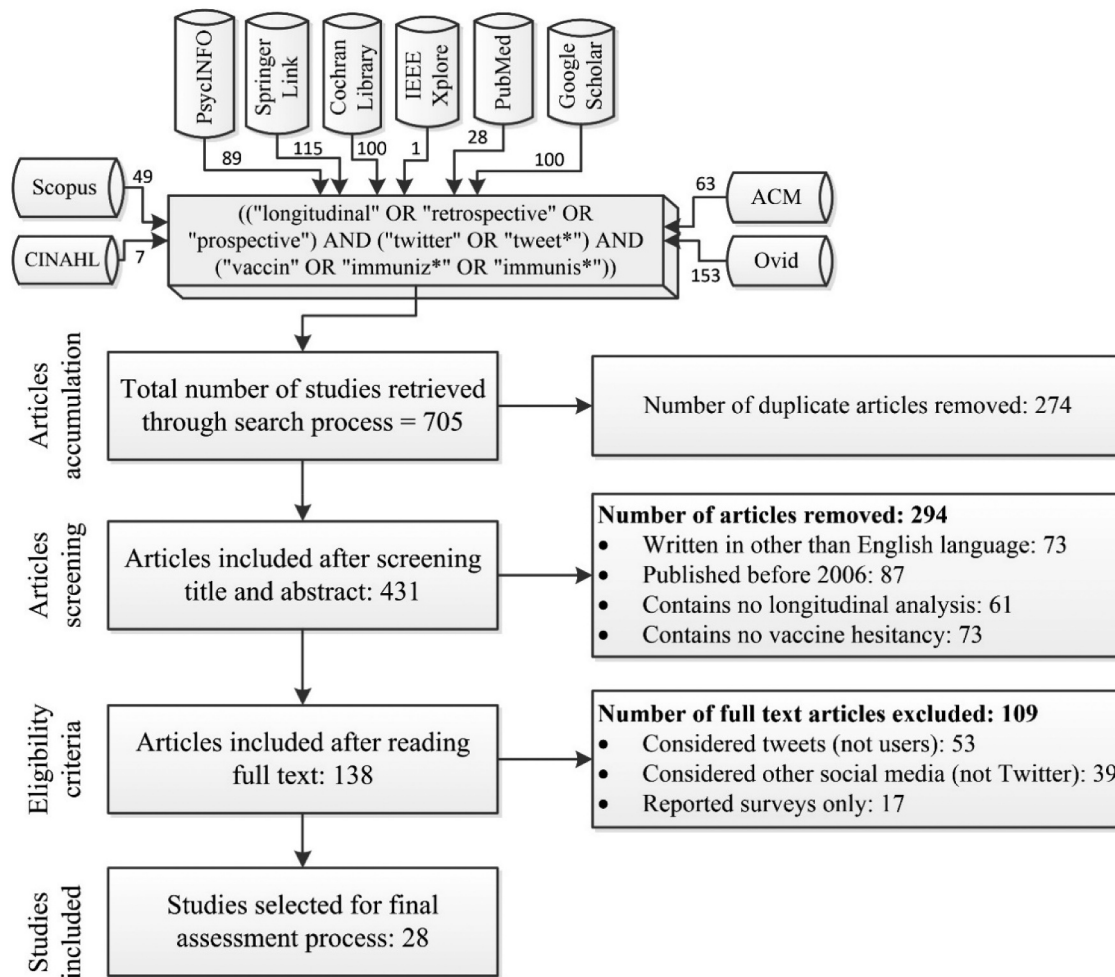


Figure 1. Experimental setup.

vaccination are most recently reported. Moreover, most of the studies are reported by the US authors ($N = 14 \sim 50\%$), the UK reported three studies ($\sim 11\%$), China reported two studies, while the rest of the countries Japan, Singapore, India, and others reported only one study on longitudinal analysis using Twitter data. Figure 2 shows that most of the studies ($N = 23 \sim 83\%$) are journal articles, while only ($N = 5 \sim 17\%$) are conference papers.

Studies taxonomy

Table 1 represents the overall taxonomy of the finalized relevant articles. It contains the information about the number of data samples, type of vaccine studied, and range of years followed for users' behavior longitudinal analysis. Also, Table 1 contains the information about the techniques or methodologies used to perform this longitudinal analysis. In the included articles along with Twitter some researchers considered data from other online social media and news sources such as Calo et al.³⁹ considered data posts and status reviews from Facebook, Instagram, Reddit, and YouTube to analyze public attitude toward HPV vaccination. While Islam et al.²² considered blogs and news reports from Google, Facebook, Fact-checking agency websites, YouTube, Fact check, and television and newspaper websites. Hussain et al.³³ considered

Facebook posts along with Twitter data for user behavior analysis toward COVID-19 vaccination.

Data source is the preliminary step for any sentiment analysis task. Numerous data sources are reported in the selected articles to identify the behavior of Twitter users toward vaccination. Among the included articles Twitter is a primary source for data accumulation in all the articles ($N = 28 \sim 100\%$) however, the research papers^{22,33,39} ($N = 3 \sim 11\%$) also used other social media sources like YouTube, Facebook, Instagram, Reddit, Fact check, and television and newspaper websites along-with Twitter for data accumulation and sentiment analysis process.

Only four studies^{19,33,34,43} ($N = 4 \sim 14\%$) provide data and implementation code for the public. Two research articles^{32,45} provided the public access link to the implementation code but didn't provide an access link to the data due to Twitter public policy restrictions. Three studies^{20,22,39} provided only the keys to the Twitter data (Tweet IDs and User IDs) used for experimentation. The research article⁴⁴ made their code publicly available but made their data private and available on request only. While the research articles^{26,37,43} made their data private and can provide the data based on a reasonable request to the corresponding authors. The rest of the ($N = 16 \sim 58\%$) of the articles made their simulations and data private. The information about

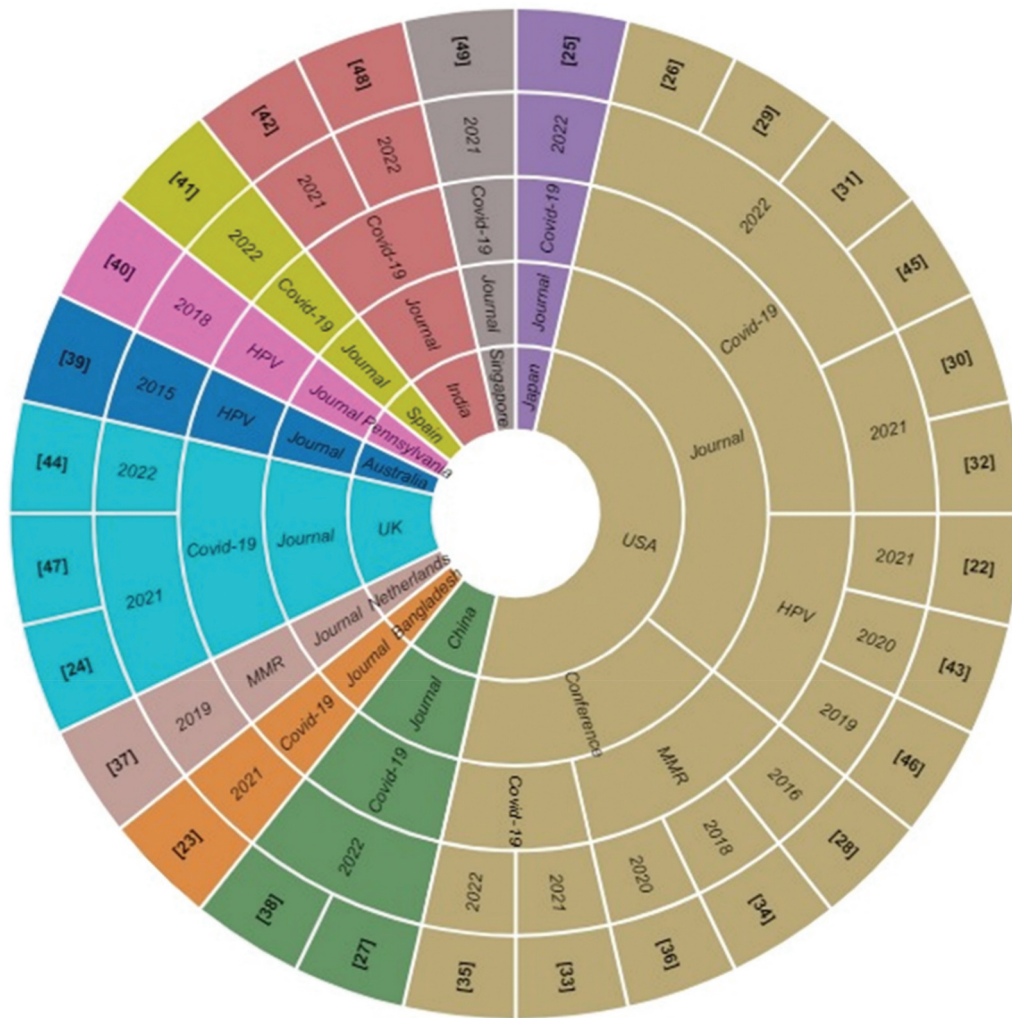


Figure 2. Evolution of included articles.

Table 2. Data and code availability with access link.

S. No	Task performed	Data access link	Code access link	Reference
1.	Monitor users' behavior	https://www.github.com/Mmichio/Aggressive_Behaviour_of_Antivaxxers_public	https://www.github.com/Mmichio/Aggressive_Behaviour_of_Antivaxxers_public	19
2.	Monitor user sentiment and emotions	Available with the paper: https://doi.org/10.1371/journal.pone.0268409	N/A	20
3.	Analyzing rumors and conspiracy theories	Available with the paper: https://doi.org/10.1371/journal.pone.0251605	N/A	22
4.	Misinformation and hesitancy calculation	Data can be available on request: (Email: tmackey@ucsd.edu)	Code can be provided on reasonable request: (Email: tmackey@ucsd.edu)	26
5.	Monitor social discussions	N/A	https://github.com/roel-sbcc/Networked-TwitterConversations	32
6.	Monitor vaccine hesitancy	https://gitlab.com/covid19aidashboard/covid-vaccination/	https://gitlab.com/covid19aidashboard/covid-vaccination/	33
7.	Public attitude toward vaccination	https://github.com/xinyuuzhou/COVID-19-vaccine-on-Twitter	https://github.com/xinyuuzhou/COVID-19-vaccine-on-Twitter	34
8.	Misinformation processing	Data available on request: (Email: iherrpec@uax.es)	Code will be provided on request: (Email: iherrpec@uax.es)	37
9.	HPV vaccine and social media	Available with the paper: https://doi.org/10.1007/s10865-021-00203-3	N/A	39
10.	Misinformation processing	https://github.com/joetien/sentinel-node-misinfo	https://github.com/joetien/sentinel-node-misinfo	42
11.	Monitor health beliefs about HPV vaccine	Data available on request: (Email: cui.tao@uth.tmc.edu)	Code will be provided on request: (Email: cui.tao@uth.tmc.edu)	43
12.	Longitudinal analysis of COVID-19 vaccine	Data can be available on request: (Email: mayur.18dr0078@cse.iitism.ac.in)	https://github.com/MayurWankhade/Sentiment-Classification-Task	45

data sources, code implementation, and public access links are shown in Table 2.

Only one research article⁴⁵ provided information about the hardware resources used for experimentation and sentiment analysis. They have used NVIDIA Tesla V100 SXM332 GB GPU for the development of their 12-layered architecture with hidden size $h = 768$, with a dropout rate of 0.2, a learning rate of 0.001, regularization of 0.001, and sigmoid as an activation function.

The finalized pool of relevant articles studies the longitudinal analysis of users' behavior about three different vaccine types. These vaccines include the MMR vaccine, HPV, and Covid-19. Most of the studies ($N = 19 \sim 68\%$) reported COVID-19 vaccination, ($N = 5 \sim 18\%$) reported HPV vaccine, and ($N = 4 \sim 14\%$) reported MMR vaccines in their longitudinal analysis using Twitter and other social media platforms.

For user demographics and profile data extraction, different techniques are reported in the included studies, such as articles^{20,21} that used m3 inference in Python for geographical information extraction. Article²¹ also used "Geopy and Pycountry" libraries, while²⁵ used DeepFace and⁴¹ used a named entity recognizer (NER) for user demographics and geographical information extraction. Several geographical regions are covered for users' psychological and behavioral analysis toward vaccination. Among the reported geographical regions USA is the highest reported region in ($N = 19 \sim 68\%$) studies.^{19,30,32,33,35,37,38,40,42} The second highest reported region for longitudinal analysis is the UK and reported in ($N = 8 \sim 29\%$) research studies.^{22,23,33,34,38,41,44,46} India and Pakistan reported in retrospective studies^{22,34,38} for users' behavior analysis toward vaccination. Australia and Brazil reported in ($N = 4 \sim 14\%$) studies^{21,22,34,38} for attitude analysis toward vaccination longitudinally.

Characteristic of data

The details about the size of data observed, the number of users analyzed for longitudinal observations, and the time frame for which the users' sentiments are analyzed are briefly explained below.

Data samples studied

In the finalized studies varying numbers of samples are used for longitudinal analysis of the users' behavior toward vaccine hesitancy. Only one article²¹ reported 1.4 billion tweets for retrospective analysis of user behavior. Similarly, a single article³⁴ reported about 13 million tweets for behavioral analysis toward vaccination.

In the final pool of relevant articles, some researchers considered other language tweets along with English language tweets, such as a research paper¹⁹ considered Japanese language tweets along with English tweets. Similarly, article²⁶ considered Spanish, Turkish, Japanese, Portuguese, German, Slovenian, and Dutch language tweets along with English language tweets for longitudinal analysis. Research paper³² followed Dutch language tweets along with English tweets, while the article³⁴ considered English tweets along with 90 different languages tweets belonging to

Southeast Asian, Eastern Mediterranean, and Western Pacific countries, including India, Indonesia, and Pakistan. The research paper³⁷ downloaded Spanish-language tweets along with English tweets for vaccine hesitancy analysis among Twitter users.

Timeframe of data

The included studies report different data timeframes for users' psychological observations and attitudes toward vaccine hesitancy. Some researchers performed their longitudinal observation on the data timeframe reported for two years, three years, and vice versa. Even some researchers reported their observations on a few months of Twitter data. A detailed description of the data timeframe is shown in Table 3.

From Table 5, it is concluded that only three articles^{23,40,43} have considered an observational period of more than three years. Similarly, only six articles^{20,21,28,34,36,45} considered data for one or more than one year for user psychological observation toward vaccination.

Number of users

During the assessment and evaluation process, we found that varying numbers of users are selected for longitudinal analysis toward vaccinations. Table 6 represents the selection of users in different research articles. From Table 1, it is concluded that only one study considered users of more than one million (1.15 million) for longitudinal behavior analysis. Only two papers considered users in the range of 500,000 to 1000,000 range. But most of the studies ($N = 10 \sim 36\%$) provided no user information during the psychological analysis toward vaccination.

Users longitudinal behavior analysis

Figure 3 represents the framework followed in this scoping review process for user longitudinal behavior analysis. It provides an encyclopedic overview of the two different themes identified during this systematic analysis of the literature. These two overarching themes include (1) factors influencing users' behavior toward vaccination, and (2) the methods/techniques used in the literature to perform a longitudinal analysis of vaccine hesitancy among Twitter users. These broad themes are further divided into sub-themes like "factors affecting user's behavior" are divided into community factors and individual factors, where the individual factors are classified into contextual factors, individual and group factors, or vaccine-specific factors. The community factors are classified into community-specific factors like politicians, religious and other influential activists, and media (news or advertisement team), as shown in the leaf nodes of Figure 3.

Table 3. Data timeframe selected for users' longitudinal sentiment analysis.

Data timeframe	References
≥ 24 months	23,40,43
12–24 months	20,21,28,34,36,45
1–12 months	19,22,24,25,27,29,33,35,37,39,41,42,44,46
1 week	26

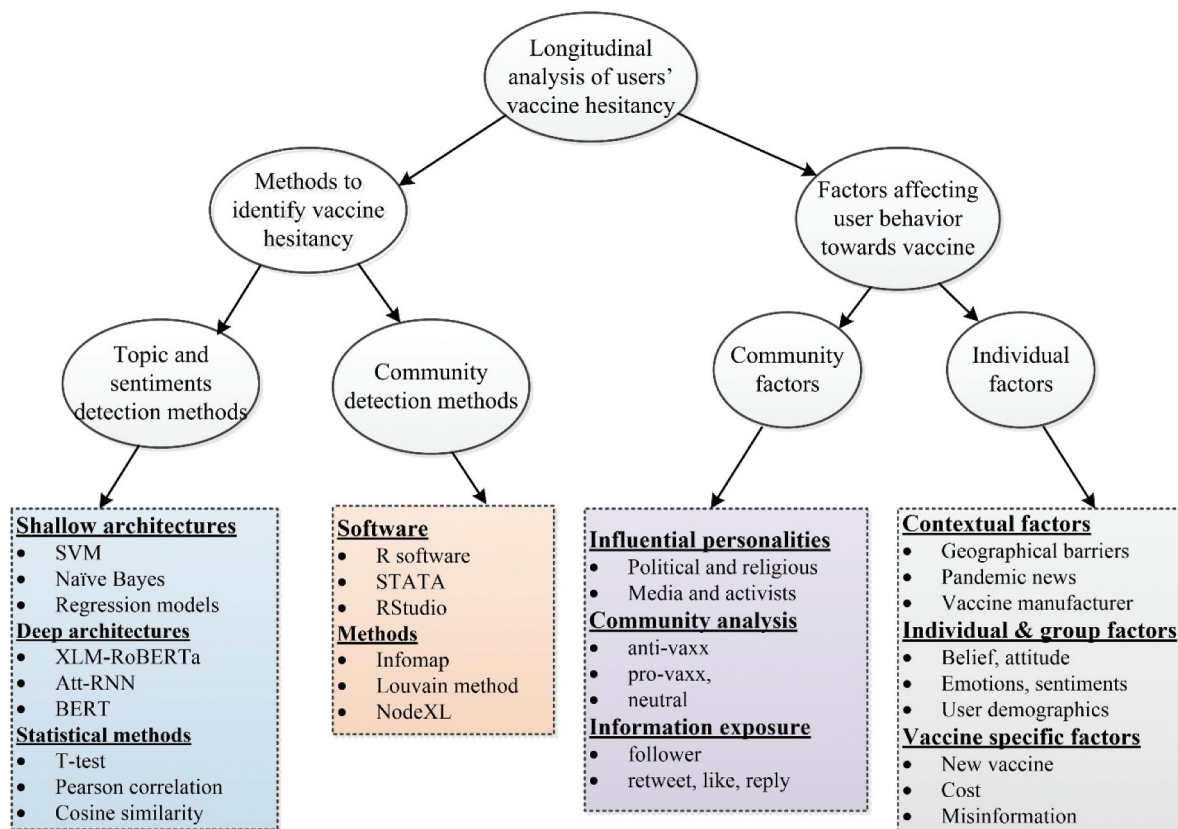


Figure 3. Study taxonomy.

The methods/techniques to identify vaccine hesitancy are divided into community detection methods that are further divided into machine learning methods and statistical methods. The machine learning methods are then further dissected into shallow and deep architectures. The shallow architectures use binary patterns for classification and identification, while the deep architectures use neural network-based models for classification and identification purposes.

Factors affecting users behavior toward vaccination

The dilemma of understanding why some people are agreeing to be vaccinated and others are not, is a critical issue, especially for the healthcare domain. The capability, opportunity, motivation – behavior (COM-B) model is presented to identify the factors influencing Twitter users' behavior toward COVID-19 vaccination.⁴⁸ The included studies identified different individual and community-based features that directly or indirectly affect public behavior toward vaccination.

Community factors

In the realm of sociology and community psychology, community factors refer to the various elements that characterize a particular community or social group. These factors encompass the collective characteristics, resources, and dynamics that influence the well-being and functioning of the community and its individual members.⁴⁹ Community factors can include aspects such as the social norms, values, and beliefs within the community, the availability of social support networks, the

quality of local institutions and services, the level of community engagement and cohesion, and the presence of economic opportunities. These factors shape the social environment and can have a profound impact on health, social relationships, and overall quality of life within a community.⁴ During this scoping review process, we identified different communities and the information disseminated from these communities that were used in the included studies for psychological analysis of Twitter users toward vaccination. Furthermore, in the included studies, different community-based factors were used to identify the engagement of a user and exposure to a certain psychological behavior toward vaccination. These factors include the information a user retweets, likes, replies, followers' networks, and the community where a user engages. Table 4 represents different community-based factors identified in the included longitudinal studies.

During our scoping review analysis, we identified that the research articles^{19,23,29,31,32,35,37,42,44} reported different communities for the longitudinal analysis of users' behavior toward vaccination. These were anti-vaxxers, pro-vaxxers, or neutral communities, and based on these communities and

Table 4. Community factors extracted from the included studies.

S. No	Community factors	References
1.	Community analysis (Anti-vax, pro-vax, neutral)	19,23,29,31,32,35,37,42,44
2.	Information exposure (followers, reply, like, retweet)	19,20,28,29,32,35,42,44
3.	Influential (political, religious, media, and activists)	19,26,27,31,33,37,38,42,43

Table 5. Individual factors extracted from the included studies.

	Factors identified	References
Individual and group factors	Sentiments and emotions	20,25,27,29,30,33,35,38,40,46
	User demographics	20,21,23,28,33,36,37,39,40
	Trust in Health system and providers	27,33,36
	Risk-benefit assessment	20,27,36,39,41,43,45
	Knowledge/awareness	20,23,27,32,33,41,43
	Belief, attitude, and prevention	20,27,32,34,40,41,43,45
	Ethnicity/Race	23,26,38
Vaccine specific factors	Personal, family experiences	27,36
	New vaccines	21,26,33,40
	Vaccination campaign	27,34,39,45,46
	Reliability of vaccine	24,26,33,46
	Misinformation	26,28,30,33,37,39,41,42
	Vaccination schedule	46
	Vaccine (side-effect)	36,38,40,43
	Cost	43
	Vaccine schedule	27,38
	Pandemic news (new cases, new deaths)	21,26,28,33,41,45,46
Contextual factors	Vaccine manufacturer	24,26,27,33,41,44
	Geographical barriers	24,26,31,38,40,43
	Culture, socio-economic	33,41,45

Table 6. Data collection strategies.

S. No	API key used	Social media platform	Articles reported
1.	Twitter search API	Twitter	19,21,25,29,31,33,35,37,40,42,43,45,46
2.	Twint (Twitter scraping tool)	Twitter	20
3.	Social scientists' team	Twitter	22
4.	Twitter Firehose stream	Twitter	23
5.	Twitter Archiver	Twitter	24
6.	Non-scientific: Crimson Hexagon Scientific: Altmetric.com API	● Audience interests' posts ● PubMed Papers ids	31
7.	set of custom scripts in Rstudio	Twitter	32
8.	CrowdTangle platform	Facebook	33
9.	Meltwater media monitoring platform	Social media monitoring and social listening platform	34
10.	Snsrape API	Twitter	38,41
11.	Manual selection of tweet by parent	Twitter	39
12.	Sprout Social—Twitter official partner platform	Twitter	44
13.	No API information	N/A	30

information exposure,^{19,20,28,29,32,35,42,44} the authors decided that a user(s) is pro-vaccine, anti-vaccine, or neutral. This information exposure is the information that a user disseminates, like what a user likes, retweets, followers, and replies. Similarly, the research articles^{19,26,27,31,33,37,38,42,43} also identified influential personalities such as political, religious, media person, and activities that a user is involved. The authors of the articles as mentioned earlier classified the user into the anti-vaccine, pro-vaccine, or neutral communities.

Individual factors

Individual factors refer to the distinct traits or characteristics that differentiate one person from another. It can be in the form of contextual factors, individual and group factors, or vaccine-specific factors.⁵⁰ Table 5 represents the individual factors identified in the included studies. Sentiment and emotions are the highly reported individual factors reported in ($N = 17 \sim 61\%$) articles.^{20,25,27,29,30,33,35,38,40,46} The highly discussed sentiments are fear, trust, sadness, anger, disgust, surprise, joy, intent to accept or reject vaccination, and many others. The second highly discussed topic in the individual and group factor is user demographics and analyzed in ($N = 13 \sim 46\%$) studies.^{20,21,23,28,33,36,37,39,40} In the user demographics, these studies reported the factors like age, gender, user type, occupation, etc. The factor belief, attitude, and prevention assessed in ($N = 9$

$\sim 32\%$) research articles.^{20,27,32,34,40,41,43,45} The highly discussed topics in this category are beliefs that vaccines are safe, effective, or non-effective, pandemic care, public safety, and many others. The individual factor knowledge/awareness is reported in ($N = 7 \sim 25\%$) studies,^{20,23,27,32,33,41,43} and the data extracted are vaccine awareness, educational impact, scientific inquiry, and many others. Detailed information about all these individual factors is provided in supplementary study taxonomy excel file.

In the vaccine-specific factors, misinformation is the highly reported theme in ($N = 11 \sim 39\%$) research articles.^{26,28,30,33,37,39,41,42} Several misinformation themes are discussed in these articles, and we divided these themes into three broad categories, including conspiracy theories, medical-related, and vaccine-related misinformation as shown in Figure 4. The research articles,^{20,23,26,32,37,41,42} reported the rumors and conspiracy theories shared on Twitter about vaccinations, while the articles^{26,30,33,34,37,42} reported their research on medical-related misinformation. The research paper^{20,23,26,33,39,42} performed their longitudinal analysis on vaccine-related misinformation reported in the Twitter data. The second highly reported theme in the vaccine-specific factor is the vaccine campaign and it is analyzed in five articles.^{27,34,39,45,46} In the campaign, the extracted text includes vaccine motivation, vaccine distribution, medical training, and many others. Detailed information about all

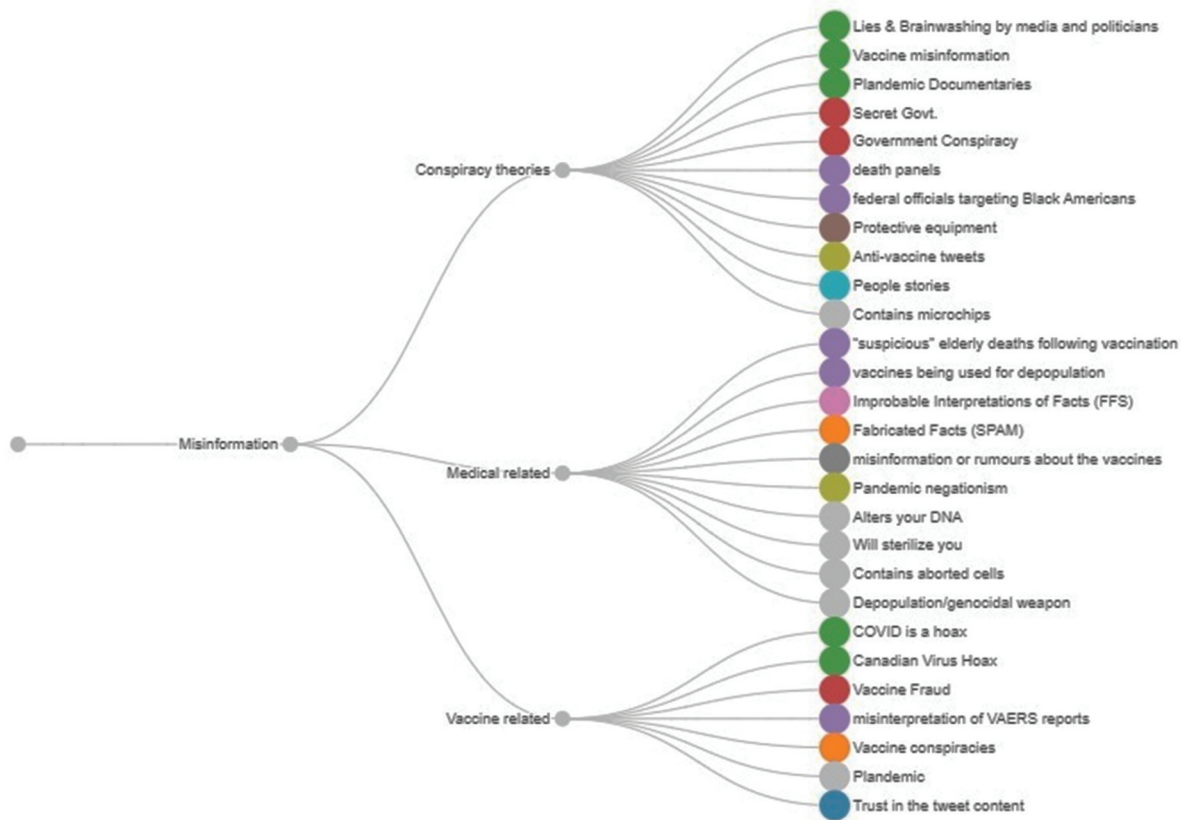


Figure 4. Misinformation analyzed in the included studies.

these individual factors is provided in supplementary material, namely, topic discussed excel file.

In psychology, contextual factors refer to the environmental or situational elements that influence an individual's thoughts, feelings, and behaviors. These factors include the physical, social, cultural, and historical contexts in which individuals are embedded. Contextual factors can have a significant impact on an individual's experiences, perceptions, and actions, shaping their development, interactions, and overall psychological well-being. During our review analysis, we identified several contextual factors in the included studies such as pandemic news reported in seven studies^{21,26,28,33,41,45,46} and data extracted relevant to this theme, including cases and deaths, reproduction rate, new case, new vaccines, vaccine & disease, and many others. Similarly, geographical barriers are also reported in seven studies,^{24,26,31,38,40,43} and the data associated with this theme include vaccine accessibility, inequities, and many others. Detailed information about all these individual factors is provided in supplementary material, namely, topic discussed excel file.

Methods to identify vaccine hesitancy

In the included studies, numerous machine learning-based methods are reported for vaccine hesitancy and users' emotional themes (happy, sad, sorrow, etc.) calculation. During our reviewing process, we found that numerous data collection strategies and statistical and machine learning-based methods are proposed for vaccine hesitancy calculation among Twitter

users. These data collection strategies and methods are briefly discussed in the following subsections.

Data collection strategies

During the analysis process, we found that different data collection strategies, such as the use of application programming interface (API) and other web-crawling keys, are used to accumulate data from social media platforms like Twitter, Facebook, YouTube, and many others. These data collection strategies are outlined in Table 6. From Table 6, it is evident that most of the articles ($N = 18 \sim 64.28\%$) have reported the Twitter search API for data accumulation from Twitter. Because the Twitter search API is provided by the Twitter developers portal and freely accessible to all the researchers globally. It is maintained by the Twitter developers' community. Two articles^{38,41} have reported the Snsrape API. It is worth noting that Snsrape extends beyond scraping tweets and offers functionality for extracting data from various other social networking platforms, including Facebook, Instagram, Reddit, VKontakte, and Weibo (Sina Weibo). Twitter permits the use of polite crawlers. Nevertheless, if the data obtained through scraping is publicly shared in an unconventional manner, Twitter has the authority to terminate API access and potentially take disciplinary measures against the account.

Community factors detection methods

The social media platforms like Twitter social media platforms that assist multinational companies, political parties, and advertising teams by proposing a dynamic perspective to

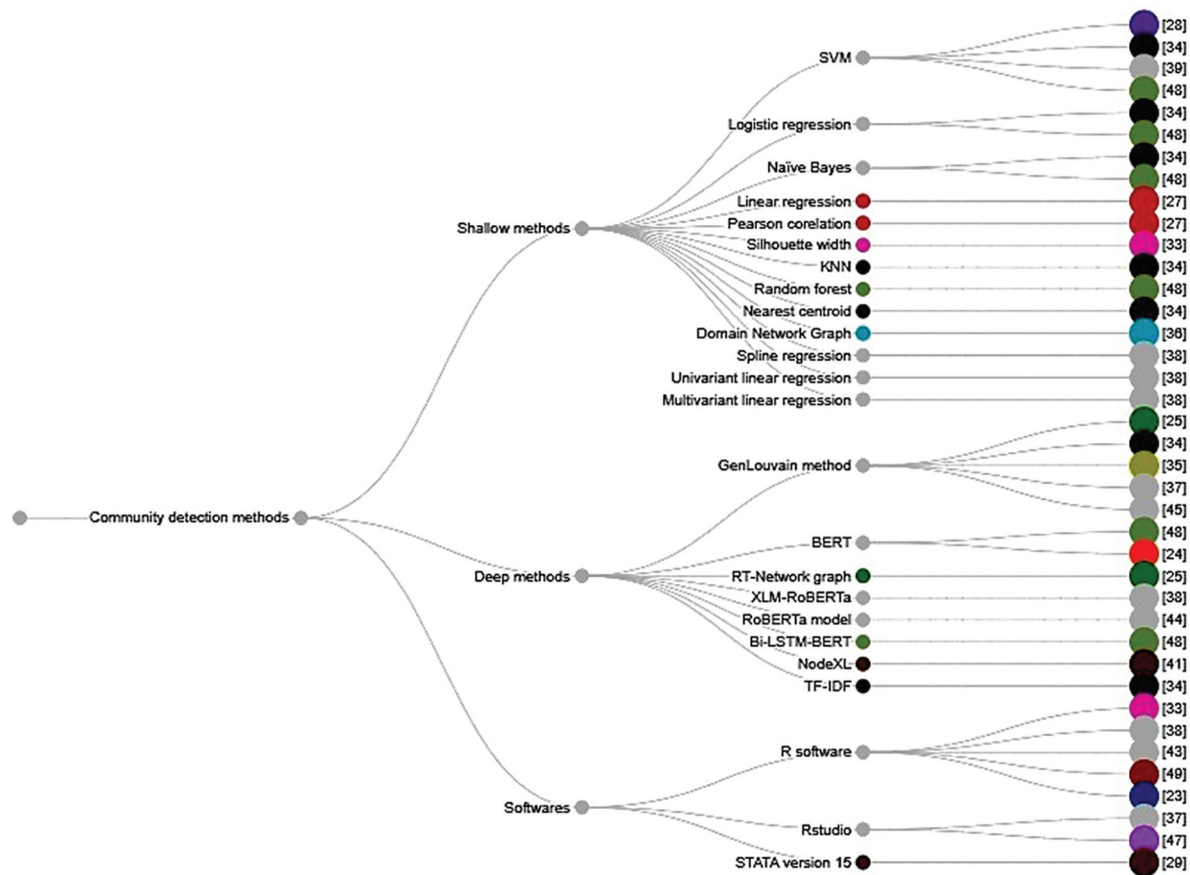


Figure 5. Different community detection methods reported in the included studies.

classify like-minded consumers and voters through community detection methods.⁵¹ In the included studies, several community detection methods are used, which are shown in Figure 5. These methods can be software, machine learning, and statistical methodologies.

Software. In the included studies, several software are reported to identify communities/clusters as shown in Figure 4. The R software is among the highest reported tools used for community detection. It is worth mentioning that different versions of R software are reported, like article²² used R version 4.0.3. Articles^{28,46} used R version 4.0.2,³⁴ used R version 4.3.2, and⁴⁰ used R version 3.6.2. Two studies^{32,44} reported the RStudio software for community detection using retweet packages. The research article²⁴ reported STATA version 15 for identifying distinct types of communities to perform longitudinal analysis of users' behavior toward COVID-19 vaccination using Twitter data.

Machine learning methods. During the assessment and evaluation process of the included studies, we found that numerous machine learning methods are proposed for users' sentiment analysis longitudinally. For simplicity, we divided these machine learning methodologies into two categories deep architectures and shallow architectures.⁵² Deep architectures are models developed using artificial intelligence and hidden layers for feature extraction, classification, and identification purposes. In the included

articles, numerous deep learning models are proposed for hesitancy analysis using Twitter data. The GenLouvain method is the most highly reported community detection method among five studies.^{19,29,30,32,42} The second highest model is BERT (Bidirectional Encoder Representations from Transformers) reported in two longitudinal studies.²² The binary invariant long short-term memory (Bi-LSTM) model is reported in⁴⁵ for longitudinal analysis of user attitude toward COVID-19 vaccination. XLM-Roberta, also known as Cross-lingual Language Model – Roberta, is reported in³⁴ for longitudinally evaluating public attitude toward vaccination using Meltwater media monitoring platform data.

The shallow architectures are techniques that have no hidden layers, and no automatic feature extraction capabilities. Some feature engineering processes are required to extract astute information from data and then perform training and testing processes of these models to accomplish identification and classification tasks. Typically, these models show outstanding performance in binary classification problems. Numerous shallow methodologies and models are reported for identifying different communities (anti-vaxxers, pro-vaxxers, or neutral) from Twitter data. In the included studies, support vector machine (SVM) is the highly utilized shallow model reported in four studies.^{23,29,35,45} Selecting an accurate kernel space, the SVM model shows outstanding capabilities in data mining and NLP-relevant research problems. Multiple regression

techniques such as logistic regression, linear regression, univariate linear regression, spline regression, and multivariate linear regression techniques are reported in the articles.^{21,29,34,45} After finely tuning the hyper-parameters these regression models show an outstanding performance for NLP-relevant tasks. Naïve Bayes technique is reported in,^{29,45} and it is considered the simplest and most generalized classification model in data mining tasks.

Individual factors detection methods

After community detection, the next step is to synthesize the techniques used in the longitudinal studies to identify the topic discussed in the communities and the individual response toward these topics. During our scoping review process, multiple techniques are reported to identify individual factors (also shown in Table 5) and their sentiments and emotions toward the discussed topics among communities. For simplicity, we divided these individual topic identification methodologies into four classes, as shown in Figure 6. These four broad classes include (1) emotion and sentiment detection techniques, (2) correlation

identification techniques, (3) topic identification techniques, and (4) user stance calculation techniques.

Sentiments and emotions detection techniques. In psychology, sentiments and emotions are related concepts, but they have some differences in their meaning and usage.⁵³ Sentiments are broader and can be shared by groups or communities. And emotions are specific to the individual and the immediate context. While emotions and sentiments are distinct, they are interconnected and can influence each other. Emotions can contribute to the formation of sentiments, and sentiments can shape emotional experiences and responses in different situations.⁵³ During our review analysis, we separated the sentiments and emotions detection techniques for readability and understandability purposes, as shown in Figure 6.

In the included studies for sentiment analysis, VADER (Valence Aware Dictionary and Sentiment Reasoner) is highly reported lexicon and rule-based sentiment evaluation tool used for sentiment analysis in the finalized longitudinal studies. It is reported in ($N = 8 \sim 28\%$) studies.^{20,21,25,27,30,33,46} A Python library TextBlob is reported in three studies.^{20,33,46}

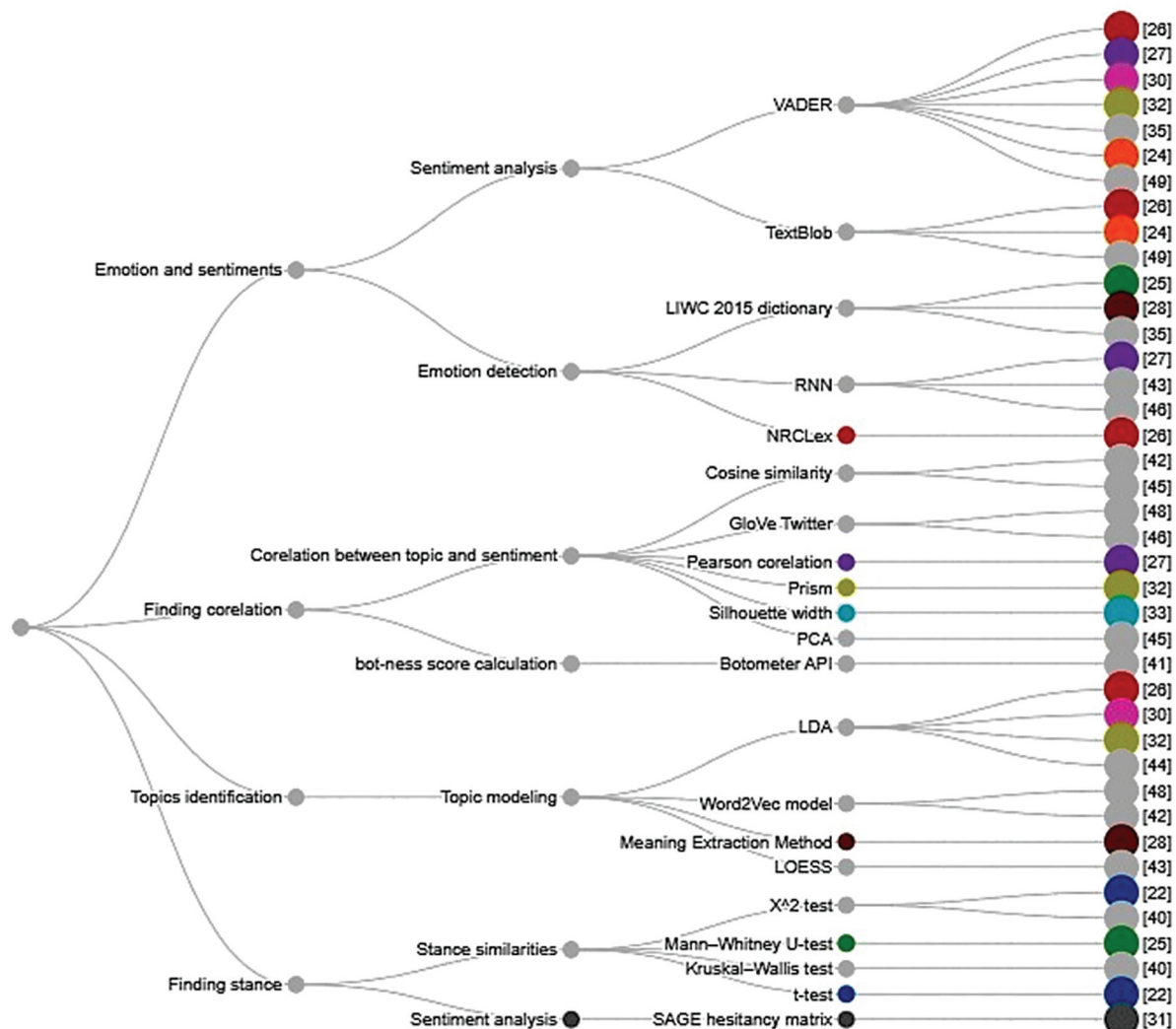


Figure 6. Individual factors and methodologies.

For emotion detection, LIWC (Linguistic Inquiry and Word Count) dictionary is used to analyze the linguistic and psychological dimensions of written text or contents in the posts. It is reported in^{19,23,30} RNN (Recurrent Neural Networks) is suggested for emotion detection in three studies.^{21,40,43} NRCLEX (National Research Council Lexicon), a Python library, is reported in²⁰ for emotion detection purposes.

Correlation identification techniques. These techniques are used to identify the correlation between the highly discussed topic and user sentiments. In the finalized articles, several correlations (between topics and sentiments) identification techniques are proposed. The research articles^{38,42} reported cosine similarity for correlation identification. The research papers^{43,45} used Global Vectors for Word Representation (GloVe) Twitter. It maps each token (i.e., word) in the text to a 200-dimension vector; pre-trained GloVe (trained on 2 billion tweets).⁵⁴ Research papers²¹ used Pearson correlation for stance calculations, while⁴² reported principle component analysis (PCA) and²⁸ used Silhouette width (ranging from -1 to 1) for stance calculation toward vaccination. The research article³⁷ used Botometer API to calculate whether a post and underlined stance are from a human or Bots (robots).

Topic identification techniques. These techniques are used to perform topic modeling (to extract topics discussed in different communities). Moreover, these techniques assisted in identifying the highly discussed topic in different communities and discussions. Four research articles^{20,25,27,41} have proposed Latent Dirichlet Allocation (LDA) technique for topic modeling. Two papers^{38,45} used the Word2Vec model for topic extraction from the community discussions²³ reported meaning extraction method, while⁴⁰ used locally estimated scatterplot smoothing (LOESS) for topic modeling purposes.

Users' stance calculation techniques. These techniques are used for calculating vaccine acceptance or rejection levels based on the topics discussed in the communities. The growing availability of digital data and large datasets has made the sentiment analysis domain more interesting, and⁴⁵ mining of texts has gained significant attention from researchers.⁵⁵ Using AI to analyze the emotions, attitudes, and opinions expressed in comments is a breakthrough that holds promise for identifying public opinions on vaccine hesitancy.⁵⁶ By categorizing opinions according to polarity (positive, negative, or neutral), emotions (such as anger and joy), or degree of agreement, sentiment analysis can provide valuable insights.⁵⁶ For user's stance calculations, the research articles^{36,39} employed χ^2 test, while the¹⁹ reported Mann-Whitney U-test for user's stance toward vaccination. Along with χ^2 test, the research article³⁶ also used Kruskal-Wallis test for user's stance calculation toward vaccination. The research paper²⁶ employed SAGE hesitancy matrix to identify user's stances toward vaccination based on the community discussion and information exposure.

Statistical methods

The statistical methods for data analysis encompass descriptive statistics and inferential statistics. Descriptive statistics employ measures like mean and median to summarize data. Inferential statistics, on the other hand, make conclusions based on data using tests like the student's t-test, z-test. Additional statistical techniques involve data sampling, central tendency, random variables, probability distributions, statistical inference, confidence intervals, and hypothesis testing. Several statistical operations and methodologies are reported in the included studies like²² employed statistical package R version 4.0.3 on an excel sheet to perform topic modeling in different community discussions. The article²⁴ reports Spearman correlation and statistical analysis for topic and sentiment analysis. The research papers^{26,27} employed the vaccine hesitancy matrix and Prism, version 9.0.2 statistical GraphPad software, respectively. Moreover, for the stance similarities and sentiment analysis numerous statistical techniques including χ^2 test, t-test, Mann-Whitney U-test, and many others provided in supplementary material, namely, methodologies employed excel file.

Discussions

This section of the paper presents a summary of the review's findings and results. It provides a concise overview of the principal outcomes, challenges encountered, and practical implications derived from this review work.

Main results

The culmination of 28 comprehensive longitudinal studies on user behavior analysis regarding vaccine hesitancy, utilizing Twitter data, highlights the pressing concern surrounding the proliferation of both community and individual factors and its correlation with vaccine hesitancy among users. This scoping review represents a groundbreaking endeavor as the first of its kind to analyze Twitter data and identify numerous influential factors at both community and individual levels that shape human behavior toward vaccines over time. Our analysis of users' behavior over time revealed two overarching themes: factors influencing human behavior toward vaccines and methodologies employed to calculate vaccine hesitancy among Twitter users. The individual factors influencing user behavior were further categorized into three distinct classes: contextual factors, individual and group factors, and vaccine-specific factors. Similarly, community factors were classified into three classes: influential personalities, community analysis, and information exposure. Methodologies utilized encompassed both machine learning methods and software-based approaches. While most of the included studies were conducted between 2021 and 2022, it is crucial to note that the findings predominantly reflect data from Europe and the United States. Consequently, there remains a notable dearth of information, particularly from African, Asian, and South American countries. Nonetheless, it is worth mentioning that Twitter emerged as the most extensively studied platform for longitudinal analysis, followed by Facebook, Reddit, and YouTube, respectively.

The growing body of recent evidence in behavior analysis utilizing Twitter data reflects the availability of new digital platforms and advanced data mining and machine learning techniques. The Gen Louvain method and SVM are the highly proposed methodologies for community detection, while LDA and Word2Vec models are the most employed algorithms for topic modeling and discussion identification among different Twitter communities. Exploring these behaviors can assist public health officials in tailoring their messages to address public health concerns and enhance healthcare delivery. In our analysis, the most frequently discussed factors are sentiments and emotions, followed by user demographics and the spread of misinformation on Twitter. We identified three broad categories of misinformation: medical-related, vaccine-related, and conspiracy theories. However, these categories are interconnected and can overlap, as skepticism toward vaccine development may be rooted in conspiratorial beliefs regarding hidden power structures and corrupt elites.

Digital data can help portray the dynamics of public health surveillance systems and allow public health professionals to pinpoint the general concerns or needs of the public during infectious disease events to create location-specific campaigns. For example, the finding that there is no association between community and individual discussions and resistive behaviors toward vaccination among Twitters users can reinforce the unfamiliarity of this population about the relationship between vaccine hesitancy and individual or community discussion on social media platforms. Several emotional and sentimental themes are identified during our evaluation process, and a number of techniques are reported to extract these sentimental and emotional themes from the users' community discussions. VADER and TextBlob are predominantly utilized libraries for sentiment analysis, while NRCLE, LIWC, and RNN are the highly employed techniques for emotions (happy, sad, sorrow, anger, and joy) calculation. Since longitudinal analysis is temporal and momentary analysis of user behavior so, RNN is the most employed emotion detection technique.

Interestingly, it is worth noting that the majority of the research conducted on Twitter engagement has focused on extended periods of time, spanning months or even years. There is a notable scarcity of studies investigating engagement on a more immediate, momentary scale. Out of the 28 finalized longitudinal studies, only a third encompassed long-term analysis, exceeding one year in duration.^{20,21,23,28,34,36,40,43,45} The long-term analysis offers enriched evidence to gain insights about the attitude patterns of the population that dissipate information about vaccines on Twitter and other social media tools. Conversely, around two-thirds of the studies were of a midterm nature, measuring engagement over several months up to one year.^{19,22,24,25,27,29,33,35,37,39,41,42,44,46} Notably, there was only a single study that employed a momentary approach, examining engagement over the course of just one week.²⁶ These momentary approaches typically employ longitudinal designs to analyze the captured data. However, further research is required to explore the short-term cross-sectional progression of engagement, specifically in relation to discussions

surrounding vaccinations at both the community and individual levels. Additionally, investigating the interaction between momentary engagement and other variables of a momentary nature is also an area that warrants additional attention.

Our research findings indicate a notable prevalence of Twitter data usage in the analysis of human behavior, particularly in the context of understanding the factors that influence user behavior toward vaccination. There are several possible explanations for this observation. Firstly, this category encompasses a wide range of topics, including seasonal outbreaks, epidemics, sexually transmitted diseases, and infectious diseases. The diverse nature of these topics makes them highly relevant and widely studied. Another contributing factor is the convenience of utilizing relative search volumes on Twitter, access logs from other social media platforms, and the prevailing fear and hype surrounding infectious diseases and various epidemics such as HPV and MMR. Surprisingly, a minimal proportion of research papers (0.3%) focused on community analysis, and a similar percentage (approximately 0.3%) explored the concept of information exposure. This finding is unexpected given the wealth of available Twitter data for analysis in these areas. A survey conducted across 19 countries between June 16 and June 20, 2020, using an online panel of 13,426 respondents, found that 72% of participants were either very or likely to take a COVID-19 vaccine. However, acceptance rates varied significantly between countries, ranging from 90% in China to less than 55% in Russia.⁵⁷ Higher vaccine acceptance was associated with older age, higher socioeconomic status, and trust in the government.⁵⁷ A recent survey of UK adults yielded similar results, with 72% of participants expressing willingness to be vaccinated and the remaining 28% reporting strong hesitancy or uncertainty.⁵⁸

During our analysis, it has been observed that vaccination plays a crucial role in the fight against the pandemic. Twitter provides a user-friendly interface where individuals can freely share their perspectives and engage in discussions on various public issues, including healthcare, politics, human rights, and personal experiences. This makes it an excellent platform for conducting opinion-based textual data analytics for various real-world applications. However, the rampant spread of misinformation related to the pandemic and vaccines through social media platforms has led the World Health Organization to coin the term "infodemic." False claims regarding negative vaccine side effects, vaccine reliability, and other individual and community factors significantly influence the behavior of Twitter users toward vaccination. These factors not only diminish the severity of outbreaks but also pose challenges for healthcare agencies and workers striving to control the spread of a particular outbreak while promoting public health through the use of vaccines and other medical resources.

Challenges and future recommendations

Based on the proposed analysis, some of the recommendations are suggested that will open new gates for the research community to explore.

- User-generated content on Twitter is often subject to bias, as it tends to reflect information that individuals feel comfortable sharing, which may not accurately represent the full range of their emotions and experiences. Among the 28 studies analyzed, no longitudinal study was found that linked the findings with users' subjective experiences, whether self-reported or not, using text, image, or video data types. Therefore, there is a significant gap in research that can identify and address content biases that impact the collection and analysis of digital data for studying vaccination-related behaviors. It is crucial to conduct studies that can accurately determine and mitigate these biases in order to enhance the reliability and validity of behavioral analyses in the context of vaccination.
- The anonymity provided by the internet allows individuals with stigmatized attributes to benefit from supportive communication on Twitter. However, the challenge of accurately determining user demographics raises unresolved questions about the population biases present among internet users with diverse cultural backgrounds or socioeconomic statuses. Demographic data for most digital platforms are not representative at a national level and tend to be skewed toward younger age groups and users with higher levels of education. Unfortunately, this important topic remains significantly underreported by the research community.
- We found that no studies assessed digital media utilization for vulnerable populations (e.g., low-income, older adults, or people with a disability) who are underrepresented on different digital platforms. Studies on detecting social bots are scarce.
- For longitudinal analysis, a considerable timeframe is required to perform an enriched analysis of different individual and community factors associated with vaccine hesitancy among Twitter users. However, in the included studies, only three studies^{23,40,43} selected a range of years greater than two years. Similarly, a considerable amount of Twitter data is required to perform a momentary qualitative analysis of user behavior, but in the included 28 articles, only one study²⁵ used a dataset greater than one million tweets.
- Among the studies included in the analysis, a mere four studies (0.14%) took into account scientific and social media platforms other than Twitter for behavioral analysis. To conduct a more comprehensive and in-depth psychological analysis of user behavior, it is essential to consider other scientific and nonscientific platforms as well. Exploring these platforms as part of future research presents a valuable challenge for further investigation.
- During this review analysis, we found that the majority of the included studies (approximately 72%) performed their longitudinal analysis on Twitter data from Europe and the United States. There is no significant contribution toward longitudinal analysis from low-income countries like Pakistan, Bangladesh, India, and many other countries. This topic requires considerable attention from the research community.
- Explainable artificial intelligence (XAI) pertains to the capability of AI systems to offer clear and transparent explanations for their decisions and actions. Within the domain of behavioral analysis, XAI plays a crucial role in improving transparency, accountability, trust, and the identification and rectification of errors. Consequently, it enhances the acceptance of AI-based methods for human psychological evaluations. However, despite the significance of explainability, the studies included in our analysis did not report any research specifically addressing the explainability aspect of various approaches utilized for longitudinal behavior analysis of Twitter users. This represents a notable gap in the current literature, emphasizing the need for future studies to explore and incorporate explainability techniques into the analysis of Twitter user behavior.

Strength and limitations

The strengths and limitations of this review analysis are discussed in brief below. It is worth mentioning that there are no magnified limitations found for this scoping review, but some minor limitations are there that threaten the validity of this work. Based on these limitations, some future recommendations are proposed that should be addressed in the near future to perform a more authentic behavior analysis of the public attitude toward vaccination using Twitter data.

Strength

The following are some of the magnified strengths and applications of the proposed scoping review research work. Firstly, to the best of our knowledge, this is the first of its kind systematic review that analyzed the longitudinal literature reported in the 10 well-reputed online repositories, namely, Scopus, PubMed, IEEE Xplore, ACM, Google Scholar, PsycINFO, Ovid, CINAHL, Springer Link, and Cochrane Library to accumulated relevant studies and perform review analysis. A considerable number of studies (705 articles) were synthesized to develop a final database of 28 most relevant longitudinal studies for the assessment and evaluation process.

Secondly, this research work presents a concise summary of the key individual and community factors from Twitter data (Facebook, Reddit, YouTube, Instagram, etc., as secondary social media tools) that influence human behavior toward vaccination longitudinally. Moreover, it also outlines the algorithms and methodologies that can be employed to perform momentary sentiment and behavior analysis using social media (Twitter and other platforms) data. Furthermore, it explains different correlation functions and statistical methodologies that can be integrated to identify the correlations between highly discussed topics in communities and conclude users' emotions and stances about vaccines (or other highly discussed topics).

Thirdly, after analyzing the literature, this research work identified the gaps in the reported extant and presented new research directions for the research community to explore. Conversely, this will not only open new directions for the researchers to explore, but it will assist the healthcare

workforce and health agencies to identify public health sentiments about outbreaks or pandemics and follow precautionary measures on a priority basis.

Limitations

The current review is a scoping review of longitudinal studies on users' behavior toward vaccine hesitancy using Twitter community discussions during the last two decades. Thus, many studies using a cross-sectional design were not included in the review, and including these studies in the review might have given a different picture of how engagement has been studied across the past two decades. For example, several studies focusing on momentary engagement using cross-sectional designs have been published during the past 20 years and were not included in this review due to their cross-sectional design. However, the present review addressed the need to review the longitudinal research on users' engagements in Twitter discussions which presents a first appraisal of the evidence base that can be further developed.

The longitudinal studies reported before 2006 were also skipped during the reviewing process because the prime concern of this scoping review is to analyze the longitudinal studies reported on Twitter data. Also, we skipped those studies that considered tweets instead of users for longitudinal analysis because the main focus of this scoping review is to identify the individual and community factors from Twitter discussions that caused a momentary shift in users' behavior toward vaccination.

For this scoping review, we considered only 10 well-reputed online repositories for longitudinal studies accumulating, namely, Scopus, PubMed, IEEE Xplore, ACM, Google Scholar, PsycINFO, Ovid, CINAHL, Springer Link, and Cochrane Library. Several journals and publishers are available that publish research work, but our main objective is to target highly peer-reviewed journals and libraries that publish medical-related research work. Moreover, the last search was performed on January 07, 2023, but the research studies report daily.

Moreover, a notable limitation arises from the weak association observed between self-reported social media usage and actual utilization, as documented in reference. A considerable portion of these investigations gathered data from Twitter, primarily because Twitter has afforded researchers access to its data, rendering it more accessible than other social media platforms. Nevertheless, it's important to note that this Twitter-centric dataset may not accurately reflect a randomly selected cross-section of the population, given that its user base predominantly comprises individuals aged 25 to 34, primarily located in the United States. Additionally, it is worth mentioning that our analysis did not encompass an evaluation of the potential influence of social media bots (automated accounts) disseminating misleading information in these studies. Furthermore, we did not delve into the role of social media algorithms in contributing to the formation of echo chambers.

During our analysis and assessment process, we considered only longitudinal studies that are reported using Twitter data. The longitudinal studies reported on medical records from

manual hospital records, healthcare centers, surveys, or other verbal discussions are skipped during our review analysis. Also, we included the studies that evaluated users' sentiments and emotions about vaccines with time, not the studies that reported effects on jobs and life standards.

Conclusion

This comprehensive scoping review investigates the patterns and trends influencing Twitter users' vaccine behavior longitudinally, as reported in the literature. The research focuses on understanding the individual and community factors that influence vaccine acceptance and refusal, exploring changes in vaccination rates, and identifying techniques used to determine vaccine behaviors. For community factors detection SVM and Gen Louvain method is the highly reported among machine learning techniques, while R is among the software significantly used for community factors detection. VADER is extensively reported for sentiment analysis while LIWC is used for emotion detection regarding vaccination. Cosine similarity and GloVe are the statistical methods increasing reported for calculating correlation between topics discussed on Twitter between different communities. LDA and Word2Vec are the techniques motley reported for topic modeling, while Whitney U test, Kruskal-Wallis test, z^2 test are the techniques used for stance calculation about vaccination among Twitter users. Community factors encompass social norms, values, beliefs, social support networks, local institutions, community engagement, and economic opportunities within a community. Individual factors include sentiment, emotions, user demographics, beliefs, attitudes, prevention, knowledge/awareness, and vaccine-specific factors such as misinformation and vaccine campaigns. The findings reveal the significance of mass media in influencing information-seeking behavior. While the majority of studies focused on Twitter, it is crucial to explore other digital platforms for a more comprehensive understanding of behavior analysis. The lack of studies reporting the explainability aspect of different approaches used for Twitter users' longitudinal behavior analysis underscores the need for further research in this area. Additionally, the demographic coverage of the studies revealed a notable dearth of information from African, Asian, and South American countries, highlighting the need for more diverse geographical representation in future studies.

Further research is needed to address gaps in community analysis and information exposure, as well as to improve the explainability of approaches used for Twitter user behavior analysis.

Note

- [a] Twitter has been renamed "X" but the word Twitter will be used in this manuscript because the literature is assessed using Twitter as the main keyword for articles accumulation.

Authors' contribution

S.K. and Z.S. contributed to the research concept and overall study design. S.K. and M.R.B. carried out the scoping review process, and Z.S. reviewed

and resolved conflicts between the two authors. Z.S. supervised the study and provided technical assistance. S.K. contributed to writing the original draft. All the authors read and approved the final manuscript.

Disclosure statement

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

Funding

Open Access funding provided by the Qatar National Library.

ORCID

Sulaiman Khan  <http://orcid.org/0000-0003-2562-3477>

Md. Rafiul Biswas  <http://orcid.org/0000-0002-5145-1990>

Zubair Shah  <http://orcid.org/0000-0001-7389-3274>

Data availability statement

The complete data can be found in this research paper. Any inquiries or requests for additional information should be directed to the corresponding author (Z.S.).

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. *Journal of the American Statistical Association*, John Wiley & Sons; 2012.
- Merchant RM, Lurie N. Social media and emergency preparedness in response to Novel Coronavirus. *JAMA*. 2020;323(20):2011–2. doi:10.1001/jama.2020.4469.
- Kaur S, Kaul P, Zadeh PM. Monitoring the dynamics of emotions during COVID-19 using Twitter data. *Procedia Comput Sci*. 2020 Jan 01;177:423–30. doi:10.1016/j.procs.2020.10.056.
- MacDonald NE. Vaccine hesitancy: Definition, scope and determinants, (in eng). *Vaccine*. 2015 Aug 14;33(34):4161–4. doi:10.1016/j.vaccine.2015.04.036.
- Johnson NF, Velásquez N, Restrepo J, Leahy R, Gabriel N, El Oud S, Zheng M, Manrique P, Wuchty S, Lupu Y. The online competition between pro- and anti-vaccination views. *Nature*. 2020 June 01;582(7811):230–3. doi:10.1038/s41586-020-2281-1.
- Cinelli M, Quattrocioni W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, Zola P, Zollo F, Scala A. The COVID-19 social media infodemic. *Sci Rep*. 2020;10(1):1–0. doi:10.1038/s41598-020-73510-5.
- Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, Zarka J, Traboulsi C, Akl E, Baddour K. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. *Cureus*. 2020 Mar 13;12(3):e7255. doi:10.7759/cureus.7255.
- Moosa AS, Wee YMS, Jaw MH, Tan QF, Tse WLD, Loke CY, Ee GLA, Ng CCD, Aau WK, Koh YLE, et al. A multidisciplinary effort to increase COVID-19 vaccination among the older adults, (in eng). *Front Public Health*. 2022;10:904161. doi:10.3389/fpubh.2022.904161.
- Brewer NT, Chapman GB, Rothman AJ, Leask J, Kempe A. Increasing vaccination: Putting psychological science into action. *Psychol Sci Public Inter*. 2017;18(3):149–207. doi:10.1177/1529100618760521.
- Bedford H, Attwell K, Danchin M, Marshall H, Corben P, Leask J. Vaccine hesitancy, refusal and access barriers: The need for clarity in terminology. *Vaccine*. 2018;36(44):6556–8. doi:10.1016/j.vaccine.2017.08.004.
- N. V. A. J. P. h. r. Committee. Assessing the state of vaccine confidence in the United States: recommendations from the National vaccine advisory committee: Approved by the National vaccine advisory committee on June 10, 2015. *Public Health Rep*. 2015;130(6):573–95. doi:10.1177/003335491513000606.
- Schmid P, Rauber D, Betsch C, Lidolt G, Denker M-L, Cowling BJ. Barriers of influenza vaccination intention and behavior—a systematic review of influenza vaccine hesitancy, 2005–2016. *PloS One*. 2017;12(1):e0170550. doi:10.1371/journal.pone.0170550.
- Alamoodi AH, Zaidan B, Al-Masawa M, Taresh SM, Noman S, Ahmaro IYY, Garfan S, Chen J, Ahmed MA, Zaidan AA, et al. Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Comput Biol Med*. 2021;139:104957. doi:10.1016/j.compbiomed.2021.104957.
- Zhao Y, Leach LS, Walsh E, Batterham PJ, Calear AL, Phillips C, Olsen A, Doan T, LaBond C, Banwell C. COVID-19 and mental health in Australia – a scoping review. *BMC Public Health*. 2022 June 15;22(1):1200. doi:10.1186/s12889-022-13527-9.
- Bussink-Voorend D, Hautvast JL, Vandeberg L, Visser O, Hulscher ME. A systematic literature review to clarify the concept of vaccine hesitancy. *Nat Hum Behav*. 2022;6(12):1634–48. doi:10.1038/s41562-022-01431-6.
- Skafle I, Nordahl-Hansen A, Quintana DS, Wynn R, Gabarron E. Misinformation about COVID-19 vaccines on social media: rapid review. *J Med Internet Res*. 2022 Aug 4;24(8):e37367. doi:10.2196/37367.
- Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MDJ, Horsley T, Weeks L, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467–73. doi:10.7326/M18-0850.
- Newman PA, Reid L, Tepjan S, Fantus S, Allan K, Nyoni T, Guta A, Williams CC. COVID-19 vaccine hesitancy among marginalized populations in the U.S. and Canada: Protocol for a scoping review. *PloS One*. 2022;17(3):e0266120. doi:10.1371/journal.pone.0266120.
- Miyazaki K, Uchiba T, Tanaka K, Sasahara K. Aggressive behaviour of anti-vaxxers and their toxic replies in English and Japanese, (in English). *Humanit Soc Sci Commun*. 2022;9(1). doi:10.1057/s41599-022-01245-x.
- Lanier HD, Diaz MI, Saleh SN, Lehmann CU, Medford RJ. Analyzing COVID-19 disinformation on Twitter using the hashtags #scamdemic and #plandemic: Retrospective study, (in English). *PloS One*. 2022;17(6):e0268409. doi:10.1371/journal.pone.0268409.
- Zhang C, Xu S, Li Z, Liu G, Dai D, Dong C. The evolution and disparities of online attitudes toward COVID-19 vaccines: year-long longitudinal and cross-sectional study, (in English). *J Med Internet Res*. 2022;24(1):e32394. doi:10.2196/32394.
- Islam MS, Kamal A-HM, Kabir A, Southern DL, Khan SH, Hasan SMM, Sarkar T, Sharmin S, Das S, Roy T, et al. COVID-19 vaccine rumors and conspiracy theories: The need for cognitive inoculation against misinformation to improve vaccine adherence (in English). *PloS One*. 2021;16(5):e0251605. doi:10.1371/journal.pone.0251605.
- Mitra T, Counts S, Pennebaker JW. Understanding anti-vaccination attitudes in social media. [Georgia Institute of Technology, United States, Microsoft Research, United States, University of Texas at Austin, United States]. AAAI Press; 2016.
- Bradford NJ, Amani B, Walker VP, Sharif MZ, Ford CL. Barely tweeting and rarely about racism: Assessing US state health department Twitter use during the COVID-19 vaccine rollout, (in eng). *Ethn Dis*. 2022;32(3):257–64. doi:10.18865/ed.32.3.257.

25. Xie Z, Wang X, Jiang Y, Chen Y, Huang S, Ma H, Li D. Public perception of COVID-19 vaccines on Twitter in the United States. *medRxiv*. 2021.
26. Calac AJ, Haupt MR, Li Z, Mackey T. Spread of COVID-19 vaccine misinformation in the ninth inning: Retrospective observational infodemic study, (in eng). *JMIR Infodemiol*. 2022 Jan;2(1):e33587. doi:10.2196/33587.
27. Margus C, Brown N, Hertelendy AJ, Safferman MR, Hart A, Ciottoni GR. "Emergency Physician Twitter use in the COVID-19 pandemic as a potential predictor of impending surge: Retrospective observational study, (in eng). *J Med Internet Res*. 2021 July 14;23(7):e28615. doi:10.2196/28615.
28. London J. Tweeting the alarm: Exploring the efficacy of Twitter as a serial transmitter during the COVID-19 pandemic. Virtual Event, Germany; 2021. doi:10.1145/3458026.3462159
29. Yuan X, Crooks AT. Examining online vaccination discussion and communities in Twitter. Copenhagen, Denmark; 2018. doi:10.1145/3217804.3217912
30. Sharevski F, Devine A, Jachim P, Pieroni E. Meaningful context, a red flag, or both? Preferences for enhanced misinformation warnings among US Twitter users. Karlsruhe, Germany; 2022. doi:10.1145/3549015.3555671.
31. Addawood A. Usage of scientific references in MMR vaccination debates on Twitter. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain; 2020. p. 971–9. doi:10.1109/ASONAM.2018.8508385.
32. Lutkenhaus R, Jansz J, Bouman M. Tailoring in the digital era: Stimulating dialogues on health topics in collaboration with social media influencers. *Digital Health*. 2019;5:1–11. doi:10.1177/2055207618821521.
33. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, Ali A, Sheikh A. Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study. *J Med Internet Res*. 2021;23(4):e26627. doi:10.2196/26627.
34. Zhou X, Zhang X, Larson HJ, de Figueiredo A, Jit M, Fodeh S, Vermund SH, Zang S, Lin L, Hou Z. Global spatiotemporal trends and determinants of COVID-19 vaccine acceptance on Twitter: A multilingual deep learning study in 135 countries and territories. *medRxiv*. 2022;p. 2022.11. 14.22282300.
35. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: An observational study. *J Med Internet Res*. 2015;17(6):e144–e144. doi:10.2196/jmir.4343.
36. Massey PM, Budenz A, Leader A, Fisher K, Klassen AC, Yom-Tov E. What drives health professionals to tweet about #HPVvaccine? Identifying strategies for effective communication. *Prev Chronic Dis*. 2018;15:E26–E26. doi:10.5888/pcd15.170320.
37. Ruiz-Núñez C, Segado-Fernández S, Jiménez-Gómez B, Hidalgo PJJ, Magdalena CSR, Pollo MDCA, Santillán-García A, Herrera-Peco I. Bots' activity on COVID-19 pro and anti-vaccination networks: Analysis of Spanish-written messages on Twitter. *Vaccines (Basel)*. 2022;10(8):1240. doi:10.3390/vaccines10081240.
38. Chopra H, Vashishtha A, Pal R, Tyagi A, Sethi T. Mining trends of COVID-19 vaccine beliefs on Twitter with Lexical Embeddings. Ithaca: Cornell University Library; 2021.
39. Calo WA, Gilkey MB, Shah PD, Dyer A-M, Margolis MA, Dailey SA, Brewer NT. Misinformation and other elements in HPV vaccine tweets: An experimental comparison, (in eng). *J Behav Med*. 2021 June;44(3):310–9. doi:10.1007/s10865-021-00203-3.
40. Du J, Luo C, Shegog R, Bian J, Cunningham RM, Boom JA, Poland GA, Chen Y, Tao C. Use of deep learning to analyze social media discussions about the human papillomavirus vaccine. *JAMA Netw Open*. 2020;3(11):e2022025–e2022025. doi:10.1001/jamanetworkopen.2020.22025.
41. Evans SL, Jones R, Alkan E, Sichman JS, Haque A, de Oliveira FB, Mougouei D. The emotional impact of COVID-19 news reporting: A longitudinal study using natural language processing. *Human Behav Emerg Technol*. 2023 Mar 09;2023:7283166. doi:10.1155/2023/7283166.
42. Osborne MT, Malloy SS, Nisbet EC, Bond RM, Tien JH. Sentinel node approach to monitoring online COVID-19 misinformation. *Sci Rep*. 2022;12(1):1–15. doi:10.1038/s41598-022-12450-8.
43. Du J, Cunningham RM, Xiang Y, Li F, Jia Y, Boom JA, Myneni S, Bian J, Luo C, Chen Y, et al. Leveraging deep learning to understand health beliefs about the human papillomavirus vaccine from social media. *npj Digit Med*. 2019 Apr 15;2(1):27. doi:10.1038/s41746-019-0102-4.
44. Fazel S, Zhang L, Javid B, Brikell I, Chang Z. Harnessing Twitter data to survey public attention and attitudes towards COVID-19 vaccines in the UK. *Sci Rep*. 2021 Dec 14;11(1):23402. doi:10.1038/s41598-021-02710-4.
45. Wankhade M, Rao ACS. Opinion analysis and aspect understanding during COVID-19 pandemic using BERT-Bi-LSTM ensemble method. *Sci Rep*. 2022 Oct 12;12(1):17095. doi:10.1038/s41598-022-21604-7.
46. Liew TM, Lee CS. Examining the Utility of social media in COVID-19 vaccination: Unsupervised learning of 672,133 Twitter posts. *JMIR Public Health Surveill*. 2021 Nov 3;7(11):e29789. doi:10.2196/29789.
47. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016 Dec 05;5(1):210. doi:10.1186/s13643-016-0384-4.
48. Liu S, Liu J. Understanding behavioral intentions toward COVID-19 vaccines: Theory-based content analysis of tweets, (in eng). *J Med Internet Res*. 2021 May 12;23(5):e28118. doi:10.2196/28118.
49. Grandjean M, Mauro A. A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts Humanit*. 2016 Dec 31;3(1):1171458. doi:10.1080/23311983.2016.1171458.
50. MacDonald NE. Vaccine hesitancy: Definition, scope and determinants. *Vaccine*. 2015 Aug 14;33(34):4161–4. doi:10.1016/j.vaccine.2015.04.036.
51. Gupta SK, Singh DP. Seed community identification framework for community detection over social media. *Arab J Sci Eng*. 2023 Feb 01;48(2):1829–43. doi:10.1007/s13369-022-07020-z.
52. Khan S, Khan HU, Nazir S. Systematic analysis of healthcare big data analytics for efficient care and disease diagnosing. *Sci Rep*. 2022 Dec 26;12(1):22377. doi:10.1038/s41598-022-26090-5.
53. Park EH, Storey VC. Emotion ontology studies: A framework for expressing feelings digitally and its application to sentiment analysis. *ACM Comput Surv*. 2023;55(9):1–38. Article 181. doi:10.1145/3555719.
54. Mitra T, Counts S, Pennebaker J. Understanding anti-vaccination attitudes in social media. *Proc Int AAAI Conf Web Soc Media*. 2021 Apr 08;10(1):269–78. doi:10.1609/icwsm.v10i1.14729.
55. Dang NC, Moreno-García MN, De la Prieta F. Sentiment analysis based on deep learning: A comparative study. *Electronics*. 2020;9(3):483. doi:10.3390/electronics9030483.
56. Piedrahita-Valdés H, Piedrahita-Castillo D, Bermejo-Higuera J, Guillem-Saiz P, Bermejo-Higuera JR, Guillem-Saiz J, Sicilia-Montalvo JA, Machío-Regidor F. Vaccine hesitancy on social media: Sentiment analysis from June 2011 to April 2019. *Vaccines*. 2021;9(1):28. doi:10.3390/vaccines9010028.
57. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, Kimball S, El-Mohandes A. A global survey of potential acceptance of a COVID-19 vaccine. *Nat Med*. 2021 Feb 01;27(2):225–8. doi:10.1038/s41591-020-1124-9.
58. Freeman D, Loe BS, Chadwick A, Vaccari C, Waite F, Rosebrock L, Jenner L, Petit A, Lewandowsky S, Vanderslott S, Innocenti S. COVID-19 vaccine hesitancy in the UK: The Oxford coronavirus explanations, attitudes, and narratives survey (oceans) II. *Psychol Med*. 2022;52(14):3127–41.