

Analyzing Sentiments and Diffusion Characteristics of COVID-19 Vaccine Misinformation Topics in Social Media: A Data Analytics Framework

Mohammad Daradkeh, University of Dubai, UAE & Yarmouk University, Jordan*

ABSTRACT

This study presents a data analytics framework that aims to analyze topics and sentiments associated with COVID-19 vaccine misinformation in social media. A total of 40,359 tweets related to COVID-19 vaccination were collected between January 2021 and March 2021. Misinformation was detected using multiple predictive machine learning models. Latent Dirichlet allocation (LDA) topic model was used to identify dominant topics in COVID-19 vaccine misinformation. Sentiment orientation of misinformation was analyzed using a lexicon-based approach. An independent-samples t-test was performed to compare the number of replies, retweets, and likes of misinformation with different sentiment orientations. Based on the data sample, the results show that COVID-19 vaccine misinformation included 21 major topics. Across all misinformation topics, the average number of replies, retweets, and likes of tweets with negative sentiment was 2.26, 2.68, and 3.29 times higher, respectively, than those with positive sentiment.

KEYWORDS

COVID-19 Vaccine, Health Belief Model (HBM), Infodemic, Misinformation, Sentiment Analysis, Social Media, Topic Modeling

1. INTRODUCTION

The rapid spread of COVID-19 has caused unprecedented health, social and economic crises worldwide (Andersen, Rambaut, Lipkin, Holmes, & Garry, 2020). The World Health Organization (WHO) officially declared the outbreak a pandemic on March 11, 2020, and the development of COVID-19 vaccines has been a major undertaking for countries and international organizations to combat the spread of this pandemic (Loomba, de Figueiredo, Piatek, de Graaf, & Larson, 2021). As of December 2020, many vaccine candidates have been shown to be effective and protective in eliciting an immune response (Mulligan et al., 2020), with preliminary analyses of phase III trials demonstrating efficacy of up to 95% (Al-Qerem & Jarab, 2021; Jackson et al., 2020). However, ongoing efforts to achieve universal vaccination have been met with a fierce and devastating onslaught of fake news and misinformation about the importance, safety, or efficacy of vaccines (Guidry et al., 2021; Loomba et al., 2021; Praveen, Ittamalla, & Deepak, 2021). While part of this misinformation is simply confusing, much of it poses serious risks and harms vaccine uptake and acceptance (Wong et al., 2021). Thus, addressing the ongoing wave of misinformation about the COVID-19 vaccine

DOI: 10.4018/IJBAN.292056

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

and developing effective strategies to alleviate its impact is of paramount importance, especially as the news cycle, dominated by the unmediated spread of misinformation, is dramatically changing the nature of how people consume and report information (Ferrara, Cresci, & Luceri, 2020; Gupta & Aluvalu, 2021; Gupta, Ramadevi, Agarwal, & Shekhar Yadav, 2021).

Misinformation is often driven by highly networked modalities in which fake news, rumors, and misleading, fabricated conspiracy theories are generated, circulated, and broadly ingested through social media venues and other outlets (Al-Zaman, 2021; Islam et al., 2020). Since the outbreak of the COVID-19 pandemic, social media platforms such as Facebook, Twitter, and Instagram have become an important tool for governments and public health organizations to disseminate important information and raise public awareness about COVID-19 vaccination campaigns (Islam et al., 2020; Loomba et al., 2021). At the same time, social media is rapidly becoming the primary channel for public expression of opinion, classification of perceptions, and expression of skepticism about the vaccine and its effectiveness. While social media platforms provide immediate access to an unlimited amount of content, they also reinforce myths, fake news, and misinformation (Cinelli et al., 2020; Wani, Agarwal, & Bours, 2021). In particular, posts quickly circulate on social media claiming that vaccines are unnecessary or even harmful (Al-Zaman, 2021; Loomba et al., 2021; Praveen et al., 2021). This negative sentiment toward vaccination can influence vaccination readiness and thus lead to a decline in vaccination coverage rates (Islam et al., 2020; Marcec, Majta, & Likic, 2021; Wong et al., 2021). Population vaccination propensity is generally not static, but responds strongly to new evidence and sentiment related to COVID-19 vaccination, as well as the perceived risk of contracting the disease (Jose et al., 2021; van der Linden, Dixon, Clarke, & Cook, 2021). Therefore, to maintain vaccination coverage beyond the threshold of herd immunity, it is necessary to monitor current public concerns and sentiments about the COVID-19 vaccine and identify factors that might lead people to change their views about vaccines and their effectiveness (Chan, Jamieson, & Albarracin, 2020; Loomba et al., 2021).

Previous research has shown that in the state of a public health emergency, misinformation circulates more rapidly in social media than factually accurate information (Jose et al., 2021; O., Gupta, & Hasibuddin, 2020; Tentolouris et al., 2021). The resulting public opinion responses not only influence the extent and speed of misinformation dissemination, but also trigger social anxiety, disrupt perceived risk and preparedness attitudes, and potentially lead to adverse decision outcomes (Bavel et al., 2020; Wang, Zhou, Zhang, Evans R, & C., 2020). A number of recently published studies have examined the impact of COVID-19 misinformation on public perceptions of the pandemic (Jang, Rempel, Roth, Carenini, & Janjua, 2021; Meo, Bukhari, Akram, Meo, & Klonoff, 2021; Wang et al., 2020), the tendency of social and political groups to believe the misinformation (Singh, Jakhar, & Pandey, 2021; F. Yin et al., 2021), and adherence to public health initiatives, including the willingness to take the COVID-19 vaccine (Al-Qerem & Jarab, 2021; Guidry et al., 2021). Social media data have also been used to identify doubts about vaccines, such as their safety and adverse effects (Jose et al., 2021; Tsao et al., 2021), as well as sentiment toward vaccines in general (F. Yin et al., 2021) or toward specific vaccines, such as vaccines from Pfizer-BioNTech, Sinopharm, and AstraZeneca (Meo et al., 2021). To our knowledge, however, there is limited insight and analysis of public voices on COVID-19 vaccination that allows us to understand the impact of misinformation on vaccine uptake and achievement of immunity in the population when countries implement this vaccination regimen. In addition, it is equally important to understand the diffusion characteristics of misinformation on social media to proactively immunize the public against misinformation about vaccination and to develop effective strategies to alleviate the negative effect of misinformation about COVID-19 vaccines (Forni et al., 2021; Singh et al., 2021; Tsao et al., 2021; Wang et al., 2020).

The main purpose of this study is to scrutinize topics and sentiments surrounding misinformation about the COVID-19 vaccine on social media. To this end, a total of 40,359 tweets about the COVID-19 vaccine were collected on social media between January 2021 and June 2021. After preprocessing the social media data, we classified the tweets into misinformation and non-misinformation using

predictive machine learning models. We then used Latent Dirichlet Allocation (LDA) topic modeling (Blei, Ng, & Jordan, 2003) to understand the most relevant misinformation topics voiced by the community and formed their viewpoints on the COVID-19 vaccine. Consistent with the Health Belief Model (HBM) (Mirzaei et al., 2021; Shahnazi et al., 2020; Zampetakis & Melas, 2021), misinformation was interpreted according to the three phases of the vaccination process: pre-vaccination, post-vaccination, and post-vaccination. We then conducted a lexicon-based sentiment analysis to understand the sentiments associated with the misinformation about the COVID-19 vaccine. Finally, we conducted paired samples t-test to compare the number of replies, retweets, and likes of misinformation tweets with different sentiment levels. By identifying the major topics and sentiments related to COVID-19 vaccine misinformation in social media, this study provides new insights into the characteristics of public responses, the spread of misinformation, and the emergence of controversy in social media. It therefore identifies ways to mitigate the negative impact of online misinformation by developing effective strategies to proactively raise public awareness of COVID-19 vaccine misinformation.

The remainder of the paper is organized as follows: Section 2 presents a review of the literature on misinformation and outlines related work that has also applied topic modeling and sentiment analysis to COVID-19 social data. Section 3 introduces the dataset and research methodology used in this study for topic discovery and sentiment analysis. Section 4 presents the misinformation topics and sentiment analysis of the detected topics from the social data sample. Section 5 discusses the results of this study. Finally, Section 6 concludes this study and highlights the main findings and future work.

2. LITERATURE REVIEW

While attention to COVID-19 vaccine misinformation is relatively recent, rumors, fake news, and conspiracy theories that reinforce myths and fabricated stories to create narratives have been studied extensively. Research on misinformation uses different concepts and meanings and addresses different factors for the diffusion of misinformation, yet they exhibit common characteristics (Kim & Kim, 2020). Allport and Postman (1947) argue that fake news is primarily publicity-seeking (albeit improbable) fabricated claims that are framed for dissemination in a particular community. Similarly, Rosnow (1980) asserts that misinformation is typically characterized as factual information that people identify as relevant without confirming its authenticity. Pennycook and Rand (2019) describe rumor as an assertion that is spread verbally and predominantly among a group of individuals, but for which there is insufficient evidence for people to believe it is true. Disinformation is often referred to as a rumor, fake news, or misinformation that people spread unintentionally or intentionally (Kim & Kim, 2020). Misinformation is typically described as the intentional production and dissemination of misleading and/or fabricated information designed to confuse and deceive the public with the intent to cause mischief or gain financial, public, or political advantage (Gupta & Aluvalu, 2019; Gupta & Kumari, 2019; Kim & Kim, 2020; O. et al., 2020; Tsao et al., 2021; van der Linden et al., 2021; Wani et al., 2021).

The influx of COVID-19 misinformation, such as fear-mongering, fake experts, and conspiracy theories, since the outbreak of the COVID-19 pandemic has piqued the interest of researchers to understand the impact on public attitudes and decision making regarding the COVID-19 vaccine, especially among a reluctant public (Cinelli et al., 2020; Odium et al., 2020; Radu, 2020; van der Linden et al., 2021). In a study to understand how Indian citizens view the COVID-19 vaccine, Praveen et al. (2021) used LDA topic modeling and found that health concerns and allergies were the top two concerns voiced by Indian citizens toward the COVID-19 vaccine. Similarly, Garcia and Berton (2021) used LDA to analyze the major topics posted on Twitter related to COVID-19. They identified 12 topics and categorized them into four broad categories, including: origin of the virus; possible causes; threat to people, society, and the economy; and possible ways to reduce exposure to the virus. Mackey et al. (2020) used the Biterm Topic Model (BTM) to identify topics in tweets related to COVID-19 indicators, tests, and vaccines. Tweets were grouped into five main categories:

Symptoms and manifestations of symptoms, symptoms of failed tests, review of poor COVID-19 detection after tests, recollection of symptoms, and whether users had ever encountered COVID-19. They also emulate information dissemination and find that both reliable and suspicious information share the same dissemination patterns.

A growing body of research has also applied sentiment analysis to social media data to identify trends and orientations in community sentiments around various COVID-19-related issues, such as social disengagement, public behavior change, and COVID-19 vaccine misinformation and suggest various strategies to mitigate their influence. In addition, studies have used social media data to identify potential associations between sentiment toward vaccines and surveillance outcomes related to vaccine-preventable interventions (Alamoodi et al., 2021; Chakraborty, Bhattacharyya, & Bag, 2020; Pano & Kashef, 2020; Wilson & Wiysonge, 2020; F. Yin et al., 2021). Studies have also examined how the flow of vaccine information is influenced by citizen engagement through social media platforms (Karafillakis et al., 2021; Odlum et al., 2020). Praveen et al. (2021) reported that 47% of social media posts discussing the COVID-19 vaccine among Indian citizens had a neutral tone, while nearly 17% of social media posts had a negative tone. Shen et al. (2021) analyzed the development of four emotions: anxiety, anger, sadness, and joy, to identify possible themes associated with these emotions. Results showed that anxiety was related to lack of COVID-19 testing and medical care. Sadness was associated with issues such as loss of friends and family members, while joy was an expression of gratitude and better health. Similarly, Chakraborty, Bhatia, et al. (2020) examined the most retweeted tweets from January 1, 2019, to March 23, 2020, and concluded that despite the high number of positive and neutral tweets, the majority of retweeted tweets expressed negative sentiments.

Social media data have also been used to identify public concerns and understand sentiment in public discussions about the efficacy and feasibility of COVID-19 vaccination (Wilson & Wiysonge, 2020), assess public knowledge and perceptions about vaccination (Karafillakis et al., 2021), identify concerns that influence vaccination intent or behavior (Guidry et al., 2021), and assess the validity of medical claims about COVID-19 vaccine on social media (Al-Qerem & Jarab, 2021; Aslam, Awan, Syed, Kashif, & Parveen, 2020; Guidry et al., 2021; Wong et al., 2021; F. Yin et al., 2021; Zampetakis & Melas, 2021). Most of these studies were based on postal or online surveys or face-to-face interviews. However, such survey studies also have disadvantages, such as long research time, limited subject population and sample size, sample representativeness, poor survey results, and researcher bias. Despite the growing interest in addressing the ongoing onslaught of COVID-19 vaccine misinformation, relatively scant studies have examined topics and sentiments toward COVID-19 vaccine misinformation using social media data and how exposure to misinformation influences people's attitudes toward vaccination and its consequences (Guidry et al., 2021). Furthermore, understanding the characteristics of the spread of misinformation on social media platforms and how it affects people's sentiment and engagement in discussions about the safety, efficacy, and side effects of the COVID-19 vaccine is necessary.

3. RESEARCH METHODOLOGY

This study presents a data analytics framework to identify topics and sentiments related to COVID-19 vaccine misinformation in social media, utilizing the Twitter platform as a data source. First, a crawler is used to elicit COVID-19 vaccine information on Twitter with a high prevalence and exposure between January 2021 and March 2021. The collected tweets were then preprocessed and classified into misinformation or non-misinformation using predictive machine learning models. To gain a clear and in-depth insight into public perceptions of the vaccine and its impact, Latent Dirichlet Allocation (LDA) topic modeling (Blei et al., 2003) was applied to COVID-19 vaccine misinformation tweets. In line with the Health Belief Model (HBM) (Shahnazi et al., 2020; Wong et al., 2021), misinformation topics were interpreted and labeled according to the three phases of the vaccination process: pre-vaccination, vaccination, and post-vaccination. Because Twitter users typically express their views

and beliefs through replies to the corresponding tweets, the textual content of the replies was analyzed using a lexicon-based approach to identify positive and negative community attitudes and underlying sentiment trends related to COVID-19 vaccine misinformation. Finally, we conducted paired samples t-test to compare the number of replies, retweets, and likes of tweets with different public sentiments toward misinformation. Figure 1 illustrates the research methodology.

3.1. Data Collection and Preprocessing

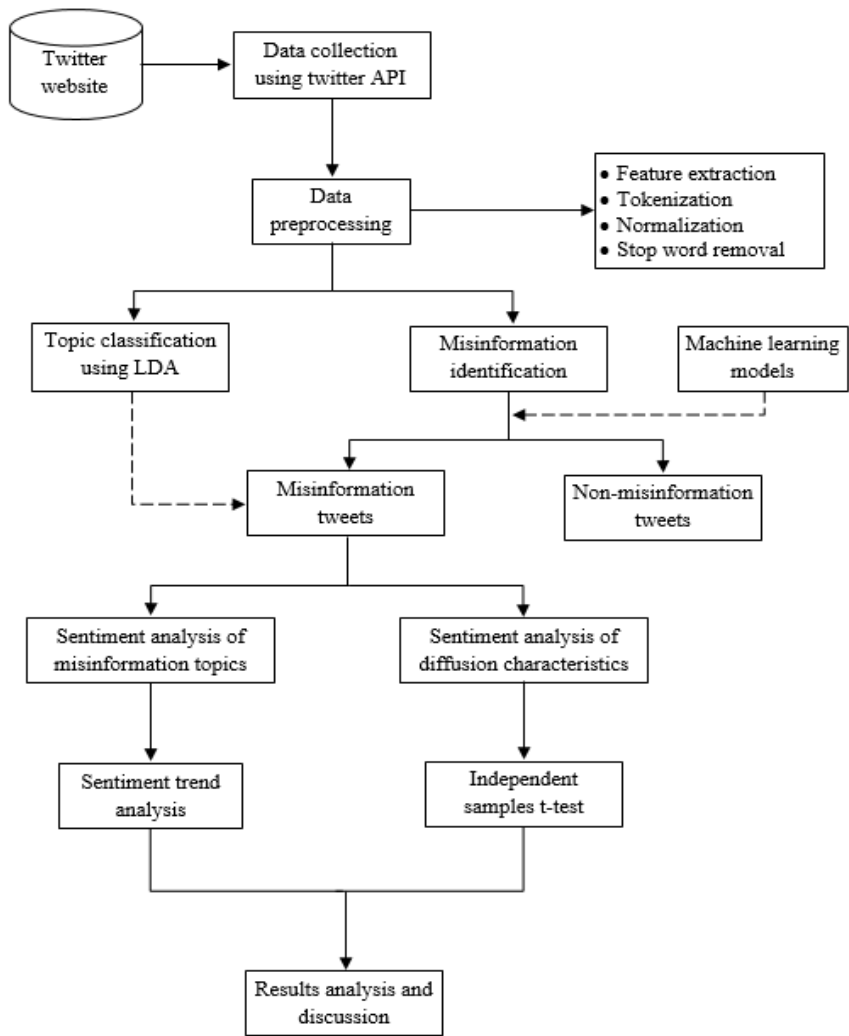
Based on the nationwide COVID-19 vaccination campaign launched by the Jordanian government in January 2021, we searched Twitter using the keyword “#vaccination” and crawled tweets with 5 or more user replies and retweets. Twitter was selected based on the amount of contributions, information, and opinions provided by public and official sources alike. Because this study targeted only English tweets, the most frequently used English keywords and hashtags (#covidvaccine, #coronavirusvaccine, #covid19vaccine, #covidvaccination, #covidvaccinefacts, #covidvaccinesideeffects, #covidvaccinearrives, #astrazenecavaccine, #vaccineefficacy, #pfizercovidvaccine, #modernavaccine, #vaccinequity, #vaccineefficacy, #vaccineschedule, #johnsonandjohnsonvaccine, #covidtreatment, and #covidvaccineupdate) were also used as alternative hashtags. These hashtags represent key issues and emerging public skepticism about COVID-19 vaccines, such as vaccine availability, vaccine side effects, acceptance, public health fears, and doubts about vaccine trials (Wong et al., 2021; Zampetakis & Melas, 2021). Tweets about vaccination for other epidemic diseases that spread across a large region, such as measles, SARS, Ebola, influenza vaccine, or other pediatric infectious diseases, were excluded. In addition, only tweets that were within the vaccine quarantine period in Jordan, which extends from January 2021 to March 2021, were collected. The content of the crawled tweets included the text of the tweets, content of replies, number of retweets, replies, and likes, number of followers, and whether the user was authenticated or not. To improve the efficiency of the estimation model, the z-score is calculated and applied to the number of retweets, replies, likes, and followers of the tweets.

In this study, a data collector was implemented using the Twitter API (version 2.0.4) and the Twitter4J library (<https://twitter4j.org/en/index.html>) to retrieve tweets along with their corresponding user profile characteristics for subsequent analysis (Gupta & Aluvalu, 2019, 2021; Gupta & Kumari, 2019). After eliminating missing unrelated and duplicate data, a total of 40,359 tweets were collected and compiled into the database. The collected raw text data was then preprocessed and cleaned using natural language processing techniques that include text normalization, stop word removal, tokenization, and case transformation. Data preprocessing is an essential step in the proposed approach, especially for social media data that is unstructured, large, noisy, and dynamic. Moreover, the textual content of tweets and replies often contains many emoticons, special characters, and web links. The primary objective of this step is to prepare the collected data for training and testing the models. In the collected data, we removed all emojis, URLs, Hypertext Markup Language (HTML), and special characters. In the case of hashtags, two cases were distinguished with respect to tweets. The first case was tweets that included hashtags as part of the sentence and added context to complement the intended meaning. Such hashtags were retained in tweets because they helped to correctly indicate the underlying meaning. In the second case, hashtags were considered an unlikely part of the tweet because they had no impact on understanding the meaning. These hashtags were discarded.

3.2 Misinformation Identification

After preprocessing the text and tokenizing it into bag-of-words vectors, the next step was to categorize the tweets into misinformation and non-misinformation. Misinformation refers to incorrect or misleading information that is considered erroneous at a given point in time according to the best available and relevant knowledge from experts. In contrast, non-misinformation refers to information that is judged to be correct by relevant experts at a given point in time based on the best available evidence (Loomba et al., 2021). In a study on rumor detection in social media, Ke et al. (2020)

Figure 1. Study methodology



showed that user credibility and microblog trustworthiness can be best characterized by including three types of attributes in the classification model: microblog attributes (including the number of retweets, likes, and replies), user attributes (including the number of followers, fans, and whether they are authenticated), and text-topic distribution. Accordingly, we classified tweets as misinformation or non-misinformation based on subsequent consideration of tweets and user attributes describing user trustworthiness and reliability. In addition, legitimate available sources of evidence, such as peer-reviewed academic literature, public health websites, and fact-checking platforms (Al-Zaman, 2021; Loomba et al., 2021), were used to confirm the authenticity of the information and the context of its presentation (Ibn Nawab, Shahiduzzaman, Eng, & Jamal, 2020).

In this study, four widely used prediction techniques were used to identify misinformation tweets, build prediction models, and compare them with each other. The techniques include decision tree (DT), K-nearest neighbors (KNN), support vector machine (SVM), and Naïve Bayes (NB). Given their predictive power, these techniques were selected based on their ability to model both a classification and a regression prediction problem and their popularity in the recent data analytics literature

(Aljameel et al., 2021; Sakr et al., 2017; Sharda, Delen, & Turban, 2021; Shen et al., 2021). We used four measures (accuracy, precision, recall, and F1 score) to evaluate the performance of these models (Sharda et al., 2021). Tweets with inconsistent prediction results from these prediction models were eliminated, while those with consistent results were retained for further analysis.

3.3 Modeling and Detection of Misinformation Topics

To reveal the main topics voiced by the community regarding COVID-19 vaccine misinformation, we used Latent Dirichlet Allocation (LDA) (Blei et al., 2003), a widely used method in the data analytics literature (Daradkeh, 2019d, 2021b; Garcia & Berton, 2021; Maier et al., 2018). LDA is a three-layer Bayesian probability model that groups frequently co-occurring words into topics. LDA treats each document as a mixture of topics and each topic as a distribution over a collection of words (Blei et al., 2003; Rohrer, Brümmer, Schmukle, Goebel, & Wagner, 2017). As a result, LDA generates two probability distributions; one is the topic-by-document matrix and the other is the word-by-topic matrix. The word-by-topic matrix indicates how a group of words can be used to form a topic. Conversely, the topic-by-document matrix reveals how people perceive important topics and in turn can indicate people's sentiment tendencies toward topics (Daradkeh, 2021b). Because the goal of this study is to analyze community sentiment toward COVID-19 vaccine misinformation topics, only relevant sentiment and most relevant COVID-19 vaccine misinformation topics were examined. The optimal number of topics was determined by calculating the perplexity of different numbers of topics (Rohrer et al., 2017). The perplexity score is a means of measuring the goodness of fit of a topic model (Blei et al., 2003). In general, the smaller the perplexity value of the model, the better it fits the topics occurring in a corpus and their relevance to different documents (Daradkeh, 2019a, 2019b, 2019c, 2019d, 2020, 2021a, 2021b, 2021c, 2022; Daradkeh & Al-Dwairi, 2018; Daradkeh & Sabbahain, 2019; Maier et al., 2018).

3.4 Sentiment Analysis of Misinformation Topics

Sentiment analysis uses natural language processing, machine learning, and text analysis techniques to systematically extract and evaluate sentiments and opinions on a given topic and classify them into predetermined, interdependent categories (Liu, 2012). These categories imply binary classifications of sentiments (positive and negative) and are usually depicted by numerical codes for subsequent statistical analyses. To date, SA techniques have reached a fairly mature stage (Chakraborty, Bhattacharyya, et al., 2020). As a result, there are currently a large number of tools and readily available packages, such as VADER (Hutto & Gilbert, 2014), SentiWordNet (Baccianella, Esuli, & Sebastiani, 2010), SentiStrength (Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010), and Word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), that are specifically designed for social media contexts and communications.

In this study, we used Valence Aware Dictionary and sEntiment Reasoner (VADER) to identify sentiment polarity from tweets and related user data. VADER provides a lexicon and rule-based sentiment analysis framework designed specifically for the social media context (Pano & Kashef, 2020). Since its introduction by Hutto and Gilbert (2014), it has been widely used for sentiment analysis tasks on various social media platforms (Amin, Hossain, Akther, & Alam, 2019; Pano & Kashef, 2020; H. Yin, Yang, & Li, 2020; Zhou, Yang, Xiao, & Chen, 2021). A main advantage of VADER is that no data preprocessing and model training are required and can be performed directly on the original tweet to induce the sentiment polarity pattern. To classify sentiment into positive, neutral, or negative categories, VADER applies a composite value calculated by summing the sentiment values associated with every term represented by the lexicon and normalizes it to return in the range (-1,1), where -1 is the most extreme negative value and 1 is the most extreme positive value. If the value exceeds 0.05, the output is positive; if it is less than -0.05, the output is negative (H. Yin et al., 2020). Accordingly, in this study, VADER was used to estimate the sentiment polarities of all original

tweets as positive or negative. The final sentiment score of the tweets was then combined with the results of the LDA topic modeling technique to analyze the sentiment orientation at the topic level.

3.5 Statistical Analysis

Paired samples t-test was applied to independent samples to compare the number of replies, retweets, and likes of tweets with positive and negative sentiment in the replies. The aim of this step is to investigate the differences in the spread characteristics of misinformation topics with different sentiments. In this study, the chi-square test was used to assess the significant difference between the sentiment distribution of responses in misinformation tweets and that of responses in non-misinformation tweets.

4. RESULTS

4.1 Misinformation Identification

In the labeling process, all collected data were searched and the tweets shared by official sources and health authorities were classified as factually correct, trustworthy, and thus non-misinformation. In the case where the relevant information could not be found through official government and health channels, or the relevant information did not exactly match the scientific content published in peer-reviewed scientific research papers and health organization websites, it was classified as misinformation.

The results of the misinformation classification prediction models are summarized in Table 1. Among the four classification methods, the C4.5 decision tree (DT) achieved better prediction accuracy. Using 10-fold cross-validation, the DT classification model showed an overall accuracy of 80.8%, followed by SVM with an accuracy of 78.2%, KNN with an accuracy of 75.9%, and NB with an accuracy of 73.9%. A paired t-test was used to determine that these accuracy results were significantly different at the 0.05 alpha level, i.e., DT is a better predictor for this field than other classification methods. Since the DT classifier showed the best performance in identifying misinformation, we then used it to classify the remaining tweets in the corpus into misinformation or non-misinformation, and that trial was done with Python. From the dataset of 40,359 tweets collected for this study, a total of 31,786 tweets were classified as misinformation and 8,573 tweets were classified as non-misinformation; this shows that false and misleading information dominates discussions on social media.

Table 1. The performance of the misinformation classification model

ML Method	Accuracy*	Precision	Recall	F-Score
DT (C4.5)	80.8	80.9	84.8	82.8
SVM	78.2	80.7	78.8	79.8
KNN	75.9	78.8	76.4	77.5
NB	73.9	77.2	73.9	75.5
* The output variable is a dichotomous variable (0 indicates non-misinformation and 1 indicates misinformation); differences were significant at $p < 0.05$.				

4.2 Misinformation Topics

The results of the LDA topic modeling for misinformation are shown in Table 2. There are two results from the LDA application, a document-topic matrix and a topic-word matrix. The topic-word matrix shows the top words in each topic. In the current study, the five most frequent words were presented for each topic. These words provide a sense of the misinformation topics surrounding COVID-19 vaccination. Specific labels for the topics can be inferred by examining these key words and the documents associated with the topics. The number of topics was set to a range between 1 and 100. The relationship between the degree of perplexity and the number of topics is shown in Figure 2. According to Figure 2, the degree of perplexity reaches the minimum value when the number of topics is 21, indicating that the model fits the data and variance best when the number of topics is 21.

As shown in Table 2, the results of this study reveal several misinformation topics that influence COVID-19 vaccine uptake, decision making, and procrastination behavior. The topic model consists of 6 dimensions: vaccination policy and strategy, society and community, organization and institution, family and friends, vaccination behavior, and vaccination experience. The vaccination policy and strategy dimension included vaccination policy, vaccination campaigns, vaccination strategy, and vaccination politicization. The society and community dimension included social health beliefs, society and community disparities, and media coverage. The organization and institution dimension included vaccine and vaccine-related organizations and vaccine availability. The family and friends dimension included family- and friendship-related factors for vaccination, such as family and friends' health beliefs, sociodemographic, previous vaccination experiences, and vaccination knowledge. The vaccination behavior dimension included vaccine efficacy, type of vaccine administered and number of doses, timing of vaccine administration, and location of vaccine administration. Finally, the vaccination experiences dimension included post-vaccination experiences, such as experiences with adverse events during vaccination, experiences with the site of vaccination administration, and experiences with vaccination personnel.

Following the Health Belief Model (HBM), we considered vaccination in this study as a process that includes pre-vaccination, post-vaccination, and post-vaccination phases. Figure 3 shows the topic model of COVID-19 vaccine misinformation. We assigned the pre-vaccination phase to the misinformation topics and dimensions that influence vaccination: individual, parents, family and friends, organization and institution, society and community, and politics. We also placed vaccination intention before vaccination. Vaccination intention included deciding whether to vaccinate in the future. Various misinformation topics influence vaccination intention, vaccination behavior, and vaccination experience. Vaccination intention influences vaccination behavior and vaccination behavior influences vaccination experience.

4.3. Sentiments and Trends of Misinformation Topics

Sentiment analysis was conducted for the classified and identified misinformation topics and related replies to misinformation tweets. The lexicon-based sentiment analysis (VADER) showed that among the misinformation tweets, there were 21,935 tweets with negative sentiment and 8,327 tweets with positive sentiment. Conversely, among non-misinformation tweets, there were 5,461 tweets with negative sentiment and 2,510 tweets with positive sentiment. Negative sentiment dominated both misinformation and non-misinformation tweets (69.32% and 64.49%, respectively), and the chi-square test for the distribution of sentiment in the misinformation and non-misinformation samples showed a statistically significant difference ($\chi^2=53.92$, $p<0.05$).

Statistical analysis of the positive and negative polarity of replies to tweets on different topics for misinformation and factually correct information was performed separately, and their polar probabilities are shown in Figure 4. Overall, the proportion of tweets with negative sentiment in the replies was higher for misinformation than for factually correct information. When comparing misinformation tweets with non-misinformation tweets, there was no significant difference in the distribution of reply sentiment for the topic of vaccine effectiveness (T14). The proportion of negative

Table 2. Topic names inferred through LDA topic modeling and the top five relevant terms (N=31,786 misinformation tweets)

Dimension/Topic Label		Top-5 Words	Freq. % M (SD)
Vaccination Policy and Strategy			
T1	Vaccination policy	Implement, vaccination, policy, approach, national	3.95 (53.90)
T2	Vaccination campaigns	Vaccination, campaign, protect, program, community	3.58 (50.37)
T3	Vaccination strategy	Prioritize, plan, immunity, exposure, strategy	2.55 (45.18)
T4	Vaccine politicization	Government, rise, skepticism, politics, politicize	2.22 (34.46)
Society and Community			
T5	Social health belief	Belief, change, behavior, health, people	4.16 (31.34)
T6	Society and community difference	Difference, community, society, compare, group	3.18 (30.40)
T7	Media coverage	Media, news, cover, report, attention,	3.57 (28.63)
Organization and Institution			
T8	Vaccination organization	Public, organization, support, promote, multiple,	2.66 (27.14)
T9	Vaccination availability	Available, provide, location, search, local	4.24 (27.98)
Family and friends			
T10	Family and friends' health belief	Belief, close, member, family, perception	4.58 (22.25)
T11	Sociodemographic aspects	Sociodemographic, factor, characteristic, difference, change	4.24 (23.90)
T12	Past vaccination experience	Symptoms, past, experience, share, receive	2.38 (16.28)
T13	Knowledge of vaccination	Knowledge, lack, share, vaccine, aware	2.38 (17.44)
Vaccination behavior			
T14	Vaccine efficacy	Efficacy, prevent, people, type, infection	4.35 (26.24)
T15	Type of vaccine	Different, type, recommend, vaccine, disease	4.76 (34.41)
T16	Vaccine dose number	Number, percentage, dose, first, total	3.18 (26.52)
T17	Vaccine administration time	Time, first, dose, administered, frame	1.55 (23.18)
T18	Vaccine administration site	Vaccination, site, administration, condition, close	1.75 (24.39)
Vaccination Experience			
T19	Experience of vaccination adverse events	Experience, side, effect, common, reaction	3.63 (35.28)
T20	Experience of vaccination administration site	Administration, site, close, place, location	3.54 (33.29)
T21	Experience of vaccination workers	Team, staff, experience, medical, train	3.21 (16.34)

sentiment in the replies for misinformation tweets was lower for vaccine dose number (T16) and vaccine administration time (T17). All other topics had a higher percentage of negative sentiment in the replies for misinformation tweets.

The sentiment trend over a six-month period (January - June 2021) on vaccine-related misinformation topics is shown in Figure 5 (months and total number of tweets with positive, negative, and neutral sentiment for each month). Negative sentiment was higher in January, February, March, and April, while positive sentiment was higher in May. There was little variation in neutral sentiment. In the first quarter of 2021, when negative sentiment increased, studies and clinical trials on the efficacy and safety of the COVID-19 vaccine were considered untrustworthy and not approved by healthcare providers and governments. In March, as negative sentiment increased, the spread of COVID-19 and associated deaths were reported. In May, as positive sentiment increased, a global initiative to support vaccination programs and rapid progress on COVID-19 vaccine for all countries

Figure 2. Perplexity of LDA model with different numbers of topics

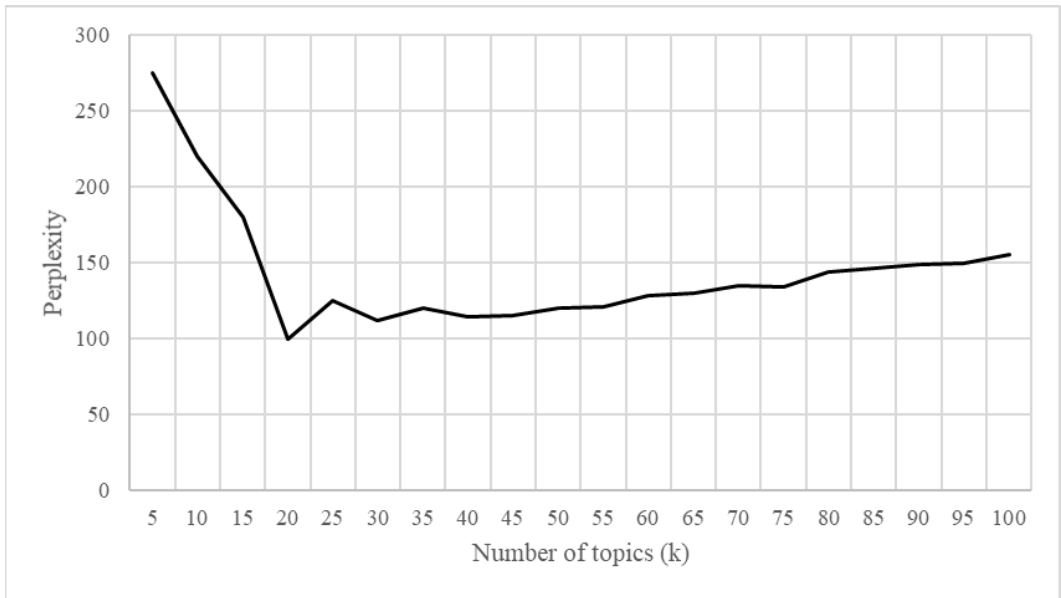


Figure 3. Conceptual framework for describing COVID-19 vaccine misinformation based on the Health Belief Model (HBM)

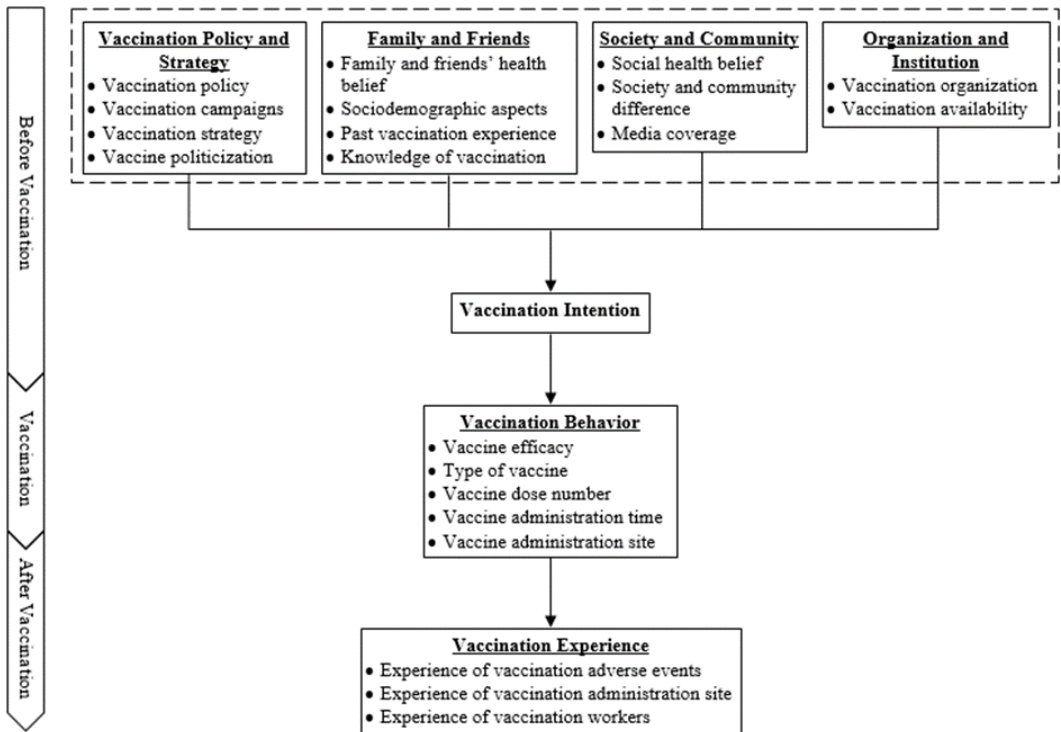


Figure 4. Sentiment proportion of each topic (N=40,359 tweets)

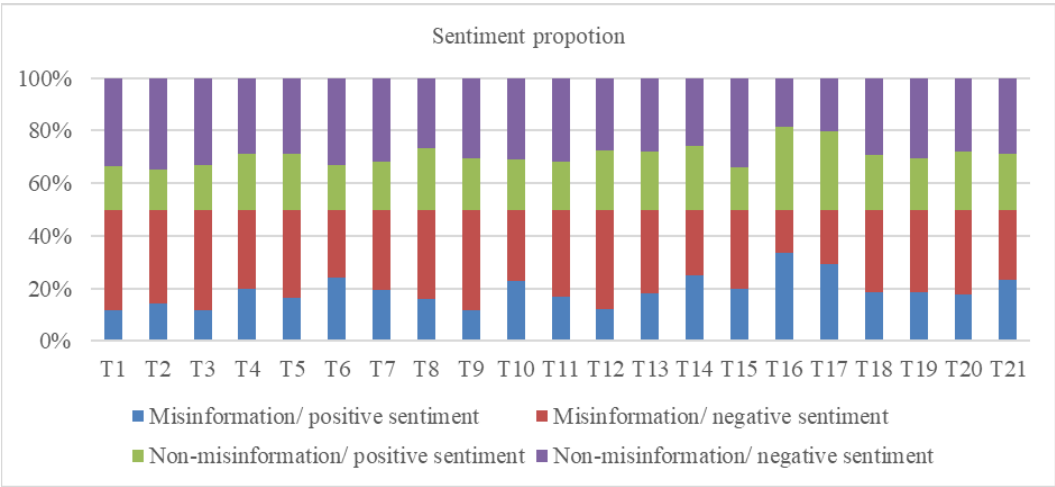
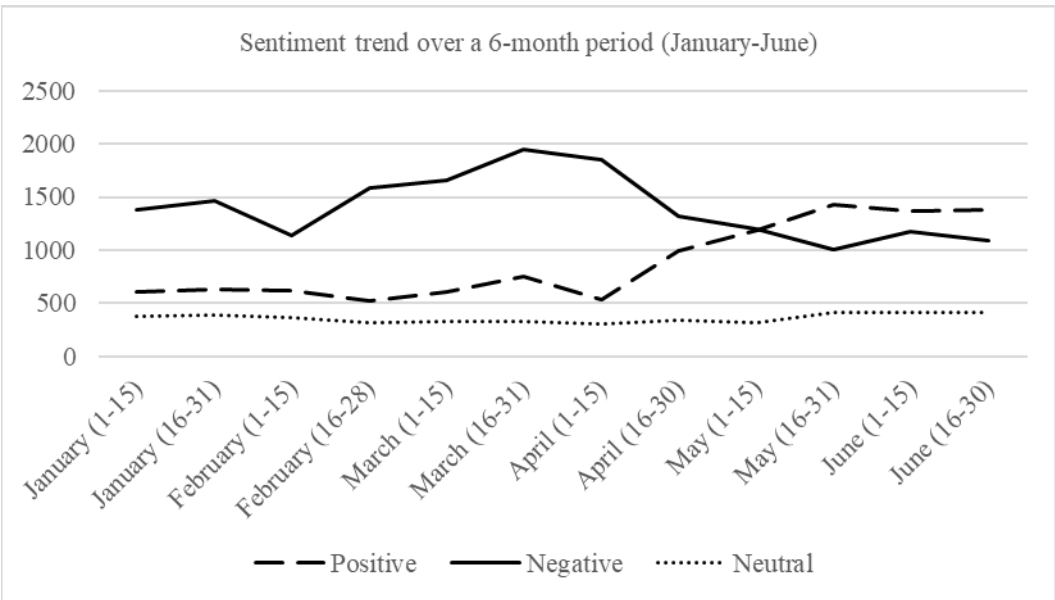


Figure 5. Sentiment trend of vaccine-related misinformation topics (misinformation tweets, N=31,786)



was reported. In May, positive sentiment also increased as WHO recommendations for COVID-19 vaccination became known.

4.4. Diffusion Characteristics of Misinformation Topics

When tweets are received, the main actions users are likely to engage in are replies, retweets, and likes, which also amplifies misinformation and spreads it widely. The results of the two independent samples t-test applied to the number of replies to misinformation tweets with positive and negative sentiment are shown in Table 3. Overall, the number of replies to tweets with negative sentiment

was higher than to those with positive sentiment, with an average of 2.26 replies. Differences in the number of replies were statistically significant for the topics of vaccination campaigns, vaccine politicization, social health beliefs, vaccination organization, family and friends' health beliefs, experiences with the place of vaccination administration, and experiences with vaccination workers. The number of replies for tweets with positive sentiment was higher for social health beliefs, type of vaccine, and vaccine administration time, while the number of replies with negative sentiment was higher for the remaining topics.

The results of the two independent samples t-test of the number of retweets with positive and negative sentiment in misinformation tweets are shown in Table 4. Overall, the number of retweets was higher for tweets with negative sentiment, averaging 2.68 retweets higher than those with positive sentiment. Differences in the number of retweets were statistically significant for the topics of social health beliefs, vaccination organization, vaccination availability, previous vaccination experience, vaccine efficacy, and vaccine administration site, with more retweets for the tweets with negative sentiment.

The results of the two independent samples t-test of the number of likes for misinformation tweets with positive and negative sentiment are shown in Table 5. Overall, tweets with negative sentiment had more likes than those with positive sentiment, with an average of 3.29 likes. Differences in the number of likes were statistically significant for the topics of vaccine availability, health beliefs of family and friends, previous vaccination experience, type of vaccine, and experience with vaccine administration site, all of which had more likes for tweets with negative sentiment than other topics.

5. DISCUSSION

This study presents a data analytics framework for analyzing topics and sentiments toward COVID-19 vaccine misinformation in social media using Twitter as a data source and study object. Despite the growing interest in studying emerging concerns around COVID-19, the identification of public attitudes and sentiments toward COVID-19 vaccine misinformation from social data is still limited. The topic model developed in this study included various factors that influence vaccination, such as individual, parent, family and friends, organization and institution, society and community, and politics. This thematic model of COVID-19 vaccine misinformation not only included topics related to COVID-19 vaccine misinformation, but also incorporated pre-vaccination intention, vaccination behavior, and post-vaccination experience. Collectively, these topics underscore the importance of managing and guiding the generation, validation, and dissemination of trustworthy information on social media (Al-Zaman, 2021; Forni et al., 2021; Jose et al., 2021).

The results of this study show that, during the COVID-19 pandemic, a significant amount of misinformation related to the vaccine exists in social media, with a high proportion of negative sentiment in misinformation compared to non-misinformation. At the same time, misinformation with negative sentiment is more likely to be re-posted and shared than misinformation with positive sentiment, with high audience engagement and interaction. According to the frequency analysis, the most common vaccination issues were social health beliefs (4.16%), availability of vaccination (4.24%), health beliefs of family and friends (4.58%), sociodemographic (4.24%), vaccine effectiveness (4.35%), and type of vaccine (4.76%). These findings have also been highlighted as important issues in other surveys of COVID-19 vaccination decisions and refusal behaviors (Al-Qerem & Jarab, 2021; Forni et al., 2021; Zampetakis & Melas, 2021). Overall, negative attitudes occurred 2.5 times more frequently than positive attitudes. Negative attitudes toward vaccination were also reported to be more common than positive attitudes in other studies based on social data (Du et al., 2020; Jose et al., 2021; Loomba et al., 2021; Praveen et al., 2021; Tsao et al., 2021).

These findings are not only consistent with previous findings of similar studies in other non-public health contexts, but also support findings reported in several studies on public health epidemics. For example, Tai and Sun (2011) used the SARS incident in 2003 and found that

Table 3. t-test of replies to tweets with positive and negative sentiment

Topic	Label	Number of replies Mean (SD)		t	p
		Pos. sentiments	Neg. sentiments		
T1	Vaccination policy	19.37 (5.22)	20.46 (16.02)	-1.06	0.29
T2	Vaccination campaigns	10.35 (11.94)	13.81 (12.87)	-2.29	0.02
T3	Vaccination strategy	18.79 (13.94)	21.27 (14.78)	-1.39	0.16
T4	Vaccine politicization	11.16 (13.35)	18.42 (15.49)	-4.18	0.00
T5	Social health belief	14.73 (9.95)	12.19 (11.53)	2.44	0.02
T6	Society and community difference	12.29 (10.64)	10.77 (10.72)	-1.89	0.06
T7	Media coverage	12.51 (12.99)	11.09 (14.28)	1.10	0.27
T8	Vaccination organization	19.12 (12.00)	22.15 (11.90)	-2.69	0.01
T9	Vaccination availability	15.22 (11.22)	14.69 (13.69)	0.19	0.85
T10	Family and friends' health belief	14.41 (12.18)	16.57 (14.33)	-4.84	0.00
T11	Sociodemographic aspects	19.37 (5.22)	20.56 (16.02)	-1.06	0.29
T12	Past vaccination experience	10.35 (11.94)	13.71 (12.87)	-2.29	0.12
T13	Knowledge of vaccination	18.79 (13.94)	21.27 (14.78)	-1.39	0.16
T14	Vaccine efficacy	11.16 (13.35)	18.42 (15.49)	-4.18	0.11
T15	Type of vaccine	14.53 (9.95)	12.19 (11.53)	2.44	0.06
T16	Vaccine dose number	11.29 (10.64)	10.87 (10.72)	-1.89	0.06
T17	Vaccine administration time	12.41 (12.99)	11.19 (14.28)	1.10	0.27
T18	Vaccine administration site	19.12 (12.00)	22.25 (11.90)	-2.69	0.11
T19	Experience of vaccination adverse events	15.52 (11.22)	14.79 (13.69)	0.19	0.85
T20	Experience of vaccination administration site	12.21 (12.18)	16.47 (14.33)	-4.84	0.00
T21	Experience of vaccination workers	17.42 (12.18)	15.45 (14.33)	-4.84	0.00
Total		12.14 (12.18)	15.77 (14.33)	-4.84	0.00
<i>Note: Statistically significant at $p < 0.05$.</i>					

relevant misinformation was widely disseminated within China, leading to public panic and social instability to some extent. Kim and Kim (2020) compared medical and health misinformation with economic, political, and military misinformation and found that the public is more likely to have strong sentiments when confronted with medical and health misinformation. This study provides a complementary and profound explanation for findings from previous studies (Baptista & Gradim, 2020; Kim & Kim, 2020). According to this study, the sentimental characteristics of misinformation and the triggered affective reactions of the public accelerate the spread of misinformation in social media. Misinformation, in turn, triggers the amplification effect of negative affective reactions in the public, leads to social panic, and influences the formulation of public policy and allocation of resources to control epidemics. Therefore, in conjunction with the results of this study, future work addressing misinformation in healthcare emergencies to control public sentiment and belief in fake news can be considered from three aspects: information source, information dissemination, and information absorption and consumption.

Table 4. t-test of retweets with positive and negative sentiment.

Topic	Label	Number of retweets Mean (SD)		<i>t</i>	<i>p</i>
		Pos. sentiments	Neg. sentiments		
T1	Vaccination policy	13.75 (10.99)	18.41 (13.87)	3.20	0.00
T2	Vaccination campaigns	15.58 (12.36)	11.63 (13.33)	-2.53	0.01
T3	Vaccination strategy	12.42 (13.75)	11.63 (15.42)	1.69	0.09
T4	Vaccine politicization	11.08 (11.21)	14.11 (12.12)	3.53	0.00
T5	Social health belief	14.76 (13.66)	20.47 (11.65)	4.19	0.00
T6	Society and community difference	10.65 (8.38)	19.20 (11.77)	-1.30	0.20
T7	Media coverage	11.49 (8.39)	10.82 (13.23)	-0.73	0.47
T8	Vaccination organization	17.33 (11.41)	21.40 (16.91)	2.92	0.00
T9	Vaccination availability	12.81 (10.35)	15.88 (14.99)	2.44	0.02
T10	Family and friends' health belief	12.35 (11.23)	16.51 (14.47)	5.94	0.00
T11	Sociodemographic aspects	19.37 (5.22)	20.56 (16.02)	-1.06	0.29
T12	Past vaccination experience	10.35 (11.94)	13.71 (12.87)	-2.29	0.02
T13	Knowledge of vaccination	18.39 (13.94)	21.27 (14.78)	-1.39	0.16
T14	Vaccine efficacy	11.16 (13.35)	18.42 (15.49)	-4.18	0.00
T15	Type of vaccine	14.73 (9.95)	12.29 (11.53)	2.44	0.02
T16	Vaccine dose number	10.29 (10.64)	10.87 (10.72)	-1.89	0.06
T17	Vaccine administration time	12.51 (12.99)	11.19 (14.28)	1.10	0.27
T18	Vaccine administration site	19.02 (12.00)	12.25 (11.90)	-2.69	0.01
T19	Experience of vaccination adverse events	15.31 (11.22)	14.79 (13.69)	0.19	0.85
T20	Experience of vaccination administration site	13.41 (12.18)	16.27 (14.33)	-4.84	0.00
T21	Experience of vaccination workers	12.41 (12.18)	16.67 (14.33)	-4.84	0.00
Total		13.43 (11.23)	16.11 (14.47)	5.94	0.00

Note: Statistically significant at $p < 0.05$.

In terms of information source, authoritative and reliable information disseminators such as government agencies, major media outlets, and key opinion leaders play a massively influential role in polarizing opinions, which can amplify (or contain) the spread of misinformation among target audiences. In the case of the COVID-19 vaccine, the government explicitly requires that epidemic data be open and transparent, that the entire health data ecosystem be shared, and that no mitigation of confirmed cases be tolerated, so the amount of misleading information in such communications is much lower compared to other information. As for misinformation identification and sentiment analysis at the information source, purely manual comparison and identification is not realistically feasible due to the sheer volume of information. In fact, both machine learning and automatic sentiment recognition methods used in this study have achieved better results reported in other studies (Jose et al., 2021; Tsao et al., 2021). Thus, this study provides compelling evidence for using both methods by governments and healthcare institutions to identify, assess, and analyze misinformation COVID-19 that can then be verified through a form of crowdsourcing to ensure higher identification accuracy and specificity.

Table 5. t-test of likes on tweets with positive and negative sentiment

Topic	Label	Number of likes Mean (SD)		t	p
		Pos. sentiments	Neg. sentiments		
T1	Vaccination policy	38.65 (37.41)	48.42 (39.10)	2.05	0.04
T2	Vaccination campaigns	41.46 (29.25)	36.24 (29.37)	-1.46	0.15
T3	Vaccination strategy	48.35 (35.77)	49.43 (36.06)	0.27	0.79
T4	Vaccine politicization	38.24 (39.17)	45.53 (39.33)	1.64	0.10
T5	Social health belief	46.42 (43.65)	45.18 (41.77)	-0.25	0.80
T6	Society and community difference	38.24 (34.55)	37.58 (30.38)	-0.09	0.93
T7	Media coverage	36.74 (33.76)	34.62 (32.63)	-0.62	0.54
T8	Vaccination organization	49.44 (46.09)	43.33 (41.92)	0.91	0.36
T9	Vaccination availability	45.15 (36.34)	56.21 (39.13)	3.05	0.00
T10	Family and friends' health belief	42.67 (38.30)	46.16 (37.74)	2.43	0.02
T11	Sociodemographic aspects	19.37 (5.22)	20.46 (16.02)	-1.06	0.29
T12	Past vaccination experience	10.35 (11.94)	13.71 (12.87)	-2.29	0.02
T13	Knowledge of vaccination	18.79 (13.94)	21.17 (14.78)	-1.39	0.16
T14	Vaccine efficacy	11.16 (13.35)	18.42 (15.49)	-4.18	0.10
T15	Type of vaccine	14.73 (9.95)	12.29 (11.53)	2.44	0.02
T16	Vaccine dose number	18.29 (10.64)	10.87 (10.72)	-1.89	0.06
T17	Vaccine administration time	12.51 (12.99)	11.19 (14.28)	1.10	0.27
T18	Vaccine administration site	19.12 (12.00)	22.25 (11.90)	-2.69	0.11
T19	Experience of vaccination adverse events	15.22 (11.22)	14.79 (13.69)	0.19	0.85
T20	Experience of vaccination administration site	14.41 (12.18)	16.67 (14.33)	-4.84	0.00
T21	Experience of vaccination workers	19.37 (5.22)	20.56 (16.02)	-1.06	0.29
Total		42.67 (38.30)	46.06 (38.30)	2.43	0.02
Note: Statistically significant at $p < 0.05$.					

In the process of information dissemination, sentiment monitoring and sentiment correction should be carried out effectively. Sentiment monitoring in public health emergencies involves quantitative and qualitative analysis of underlying trends in public sentiment. Existing literature indicates that most sentiment remediations use the fact-checking method, i.e., discouraging the audience from believing the identified misleading information through fact-checking, which influences the inference process and the emotional response of the audience (Alamoodi et al., 2021; Karafillakis et al., 2021; Radu, 2020).

In terms of information consumption and application, if the misinformation has already had a negative impact on audience sentiment, the process of responsive audience sentiment steering should be carried out, i.e., the outwardly expressed sentiment should be corrected if the sentiment has already been perceived by the audience (Loomba et al., 2021; Tsao et al., 2021). In general, detection and sentiment analysis of misleading messages, along with message correction and audience sentiment analysis, form the basis for determining the effectiveness of subsequent sentiment steering strategies.

6. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Several limitations exist in this study that should be considered when interpreting the results. First, because the intent of this study is to analyze audience sentiment and the communication characteristics of misleading information, the communication characteristics of non-misleading information were not analyzed and compared to the communication characteristics of misleading information. Therefore, we cannot rule out the possibility that the communication characteristics of non-misleading information and misleading information are similar.

Second, this study has focused on a single case study; that is, the COVID-19 outbreak. Although it met all the typical characteristics of a public health emergency, the outbreak was the most contagious and widespread of all public health emergencies in recent decades, and WHO has declared it a pandemic. Future studies could be conducted by selecting multiple outbreaks for cross-sectional comparative analysis to identify similarities, differences, and patterns among outbreaks.

Third, due to the short period of data collection, we were unable to identify annual trends in sentiment and vaccination issues. Thus, we suggest that future studies examine annual sentiment trends using data collected over a period longer than one year. We could not examine vaccination intention or behavior because there are no longitudinal studies that address these topics. In future research, different data sources, such as survey data and recent immunization reports from immunization registries, can be combined to examine immunization intent and behavior. We only used the vaccination topics classified in relation to the level of the topic model developed in this study. Therefore, in future studies, we propose to use lower-level classification categories and their relationships, such as different types of vaccinations and adverse events.

7. CONCLUSION

This study presented a systematic analysis of the main topics voiced by the public and sentiments toward COVID-19 vaccination misinformation in social media. Among the misinformation topics identified in this study, parental health beliefs, vaccination availability, and vaccination policies were identified as the most important factors associated with negative sentiment. Health beliefs can be influenced by anti-vaccine arguments, such as perceptions that natural immunity is inherently superior to vaccine-acquired immunity, or unsubstantiated rumors claiming that vaccines cause adverse reactions, illness, and even death. Therefore, it is important to monitor anti-vaccine arguments and rumors posted on social media, as they can reinforce negative sentiment toward vaccination. Vaccination availability, including cost and distance to the vaccination site, is related to vaccination policies, such as the increasing number of free vaccines and the number of health facilities offering free vaccinations. Because negative sentiments toward vaccination influence people's intention to vaccinate and lead to a decline in vaccination rates (Loomba et al., 2021; Praveen et al., 2021), it is important to find ways to improve positive sentiments toward vaccination. We anticipate that the findings from this study will serve as an indispensable reference source for practitioners and researchers to identify public concerns related to COVID-19 vaccination through social media crowdsourced data from trusted information sources and to strengthen society's resilience to misinformation about COVID-19 vaccines.

REFERENCES

- Al-Qerem, W., & Jarab, A. (2021). COVID-19 Vaccination Acceptance and Its Associated Factors Among a Middle Eastern Population. *Frontiers in Public Health*, 9(1), 632914–632914. doi:10.3389/fpubh.2021.632914 PMID:33643995
- Al-Zaman, M. (2021). COVID-19-related online misinformation in Bangladesh. *Journal of Health Research*, 10.1108/JHR-09-2020-0414
- Alamoodi, A., Zaidan, B., Zaidan, A., Albahri, O., Mohammed, K., Malik, R., Almahdi, E. M., Chyad, M. A., Tareq, Z., Albahri, A. S., Hameed, H., & Alaa, M. (2021). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, 167(1), 114155. doi:10.1016/j.eswa.2020.114155 PMID:33139966
- Aljameel, S., Alabbad, D., Alzahrani, N., Alqarni, S., Alamoudi, F., Babili, L., Aljaafary, S. K., & Alshamrani, F. (2021). A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia. *International Journal of Environmental Research and Public Health*, 18(1), 218. doi:10.3390/ijerph18010218 PMID:33396713
- Allport, G., & Postman, L. (1947). Allport, G. W. and Postman, L. The psychology of rumor. New York: Henry Holt, 1947, pp. 247. \$2.60. *Journal of Clinical Psychology*, 3(4), 402–402. doi:10.1002/1097-4679(194710)3:4<402::AID-JCLP2270030421>3.0.CO;2-T
- Amin, A., Hossain, I., Akther, A., & Alam, K. (2019). *Bengali VADER: A Sentiment Analysis Approach Using Modified VADER*. Paper presented at the 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). doi:10.1109/ECACE.2019.8679144
- Andersen, K., Rambaut, A., Lipkin, W., Holmes, E., & Garry, R. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, 26(4), 450–452. doi:10.1038/s41591-020-0820-9 PMID:32284615
- Aslam, F., Awan, T., Syed, J., Kashif, A., & Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications*, 7(1), 23–32. doi:10.1057/s41599-020-0523-3
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
- Baptista, J., & Gradim, A. (2020). Understanding Fake News Consumption: A Review. *Social Sciences*, 9(10), 185–201. doi:10.3390/socsci9100185
- Bavel, J., Baicker, K., Boggio, P., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., & Willer, R. et al. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5), 460–471. doi:10.1038/s41562-020-0884-z PMID:32355299
- Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(1), 993–1022.
- Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. (2020). Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers-A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*, 97(1), 106754–106754. doi:10.1016/j.asoc.2020.106754 PMID:33013254
- Chakraborty, K., Bhattacharyya, S., & Bag, R. (2020). A Survey of Sentiment Analysis from Social Media Data. *IEEE Transactions on Computational Social Systems*, 7(2), 450–464. doi:10.1109/TCSS.2019.2956957
- Chan, M., Jamieson, K., & Albarracin, D. (2020). Prospective associations of regional social media messages with attitudes and actual vaccination: A big data and survey study of the influenza vaccine in the United States. *Vaccine*, 38(40), 6236–6247. doi:10.1016/j.vaccine.2020.07.054 PMID:32792251
- Cinelli, M., Quattrocioni, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10(1), 16598. doi:10.1038/s41598-020-73510-5 PMID:33024152

- Daradkeh, M. (2019a). Critical Success Factors of Enterprise Data Analytics and Visualization Ecosystem: An Interview Study. *International Journal of Information Technology Project Management*, 10(3), 34–55. doi:10.4018/IJITPM.2019070103
- Daradkeh, M. (2019b). Determinants of Self-Service Analytics Adoption Intention: The Effect of Task-Technology Fit, Compatibility, and User Empowerment. *Journal of Organizational and End User Computing*, 31(4), 19–45. doi:10.4018/JOEUC.2019100102
- Daradkeh, M. (2019c). Determinants of visual analytics adoption in organizations: Knowledge discovery through content analysis of online evaluation reviews. *Information Technology & People*, 32(3), 668–695. doi:10.1108/ITP-10-2017-0359
- Daradkeh, M. (2019d). Understanding the Factors Affecting the Adoption of Project Portfolio Management Software Through Topic Modeling of Online Software Reviews. *International Journal of Information Technology Project Management*, 10(3), 91–114. doi:10.4018/IJITPM.2019070106
- Daradkeh, M. (2020). Driving Business Innovation and Sustainability through Internet of Things and Data Analytics: An Innovation Stage Model. *International Journal of Innovation. Creativity and Change*, 14(9), 149–179.
- Daradkeh, M. (2021a). An Empirical Examination of the Relationship Between Data Storytelling Competency and Business Performance: The Mediating Role of Decision-Making Quality. *Journal of Organizational and End User Computing*, 33(5), 42–73. doi:10.4018/JOEUC.20210901.oa3
- Daradkeh, M. (2021b). Exploring the Usefulness of User-Generated Content for Business Intelligence in Innovation: Empirical Evidence From an Online Open Innovation Community. *International Journal of Enterprise Information Systems*, 17(2), 44–70. doi:10.4018/IJEIS.2021040103
- Daradkeh, M. (2021c). The Influence of Sentiment Orientation in Open Innovation Communities: Empirical Evidence from a Business Analytics Community. *Journal of Information & Knowledge Management*, 20(3), 2150029. doi:10.1142/S0219649221500313
- Daradkeh, M. (2022). Organizational Adoption of Sentiment Analytics in Social Media Networks: Insights from a Systematic Literature Review. *International Journal of Information Technologies and Systems Approach*, 15(1), 15–45.
- Daradkeh, M., & Al-Dwairi, R. (2018). Self-Service Business Intelligence Adoption in Business Enterprises: The Effects of Information Quality, System Quality, and Analysis Quality. In A. Information Resources Management (Ed.), *Operations and Service Management: Concepts, Methodologies, Tools, and Applications* (pp. 1096–1118). Hershey, PA: IGI Global.
- Daradkeh, M., & Sabbahain, H. (2019). Factors Influencing the Adoption of Mobile Application Development Platforms: A Qualitative Content Analysis of Developers' Online Reviews. *International Journal of Enterprise Information Systems*, 15(4), 43–59. doi:10.4018/IJEIS.2019100103
- Du, J., Luo, C., Shegog, R., Bian, J., Cunningham, R., Boom, J., & Tao, C. et al. (2020). Use of Deep Learning to Analyze Social Media Discussions About the Human Papillomavirus Vaccine. *JAMA Network Open*, 3(11), e2022025–e2022025. doi:10.1001/jamanetworkopen.2020.22025
- Ferrara, E., Cresci, S., & Luceri, L. (2020). Misinformation, manipulation, and abuse on social media in the era of COVID-19. *Journal of Computational Social Science*, 3(2), 271–277. doi:10.1007/s42001-020-00094-5
- Forni, G., Mantovani, A., Forni, G., Mantovani, A., Moretta, L., Rappuoli, R., & Vineis, P. et al. (2021). COVID-19 vaccines: Where we stand and challenges ahead. *Cell Death and Differentiation*, 28(2), 626–639. doi:10.1038/s41418-020-00720-9
- Garcia, K., & Berton, L. (2021). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing*, 101(1), 107057. doi:10.1016/j.asoc.2020.107057
- Guidry, J., Laestadius, L., Vraga, E., Miller, C., Perrin, P., Burton, C., & Carlyle, K. et al. (2021). Willingness to get the COVID-19 vaccine with and without emergency use authorization. *American Journal of Infection Control*, 49(2), 137–142. doi:10.1016/j.ajic.2020.11.018
- Gupta, S., & Aluvalu, R. (2019). Twitter Based Capital Market Analysis Using Cloud Statistics. *International Journal of Sociotechnology and Knowledge Development*, 11(2), 54–60. doi:10.4018/IJSKD.2019040104

Gupta, S., & Aluvalu, R. (2021). Pre-Processed Tweets for Secure Capital Market Analysis Using Cloud. *International Journal of Sociotechnology and Knowledge Development*, 13(1), 1–7. doi:10.4018/IJSKD.2021010101

Gupta, S., & Kumari, N. (2019). Security Mechanism for Twitter Data Using Cassandra in Cloud. *International Journal of Distributed and Cloud Computing*, 7(2), 1–6.

Gupta, S., Ramadevi, Y., Agarwal, K., & Shekhar Yadav, C. (2021). COVID pandemic analysis using Auto-Regression-Based moving average method. *Materials Today: Proceedings*. Advance online publication. doi:10.1016/j.matpr.2021.1001.1710

Hutto, C., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text*. Paper presented at the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM), Ann Arbor, MI.

Ibn Nawab, M., Shahiduzzaman, K., Eng, T., & Jamal, M. (2020). Rumor Detection in Social Media with User Information Protection. *European Journal of Electrical Engineering and Computer Science*, 4(4), 1–8.

Islam, M., Sarkar, T., Khan, S., Mostofa Kamal, A., Hasan, S., Kabir, A., & Seale, H. et al. (2020). COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. *The American Journal of Tropical Medicine and Hygiene*, 103(4), 1621–1629. doi:10.4269/ajtmh.20-0812

Jackson, L., Anderson, E., Roupahel, N., Roberts, P., Makhene, M., Coler, R., & Beigel, J. et al. (2020). An mRNA Vaccine against SARS-CoV-2—Preliminary Report. *The New England Journal of Medicine*, 383(20), 1920–1931. doi:10.1056/NEJMoa2022483

Jang, H., Rempel, E., Roth, D., Carenini, G., & Janjua, N. (2021). Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis. *Journal of Medical Internet Research*, 23(2), e25431. doi:10.2196/25431

Jose, R., Narendran, M., Bindu, A., Beevi, N. L. M., & Benny, P. (2021). Public perception and preparedness for the pandemic COVID 19: A Health Belief Model approach. *Clinical Epidemiology and Global Health*, 9(1), 41–46.

Karafilakis, E., Martin, S., Simas, C., Olsson, K., Takacs, J., Dada, S., & Larson, H. (2021). Methods for Social Media Monitoring Related to Vaccination: Systematic Scoping Review. *JMIR Public Health and Surveillance*, 7(2), e17149. doi:10.2196/17149

Ke, Z., Li, Z., Zhou, C., Sheng, J., Silamu, W., & Guo, Q. (2020). Rumor Detection on Social Media via Fused Semantic Information and a Propagation Heterogeneous Graph. *Symmetry*, 12(11), 1806.

Kim, S., & Kim, S. (2020). The Crisis of Public Health and Infodemic: Analyzing Belief Structure of Fake News about COVID-19 Pandemic. *Sustainability*, 12(23), 9904.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. doi:10.2200/S00416ED1V01Y201204HLT016

Loomba, S., de Figueiredo, A., Piatek, S., de Graaf, K., & Larson, H. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, 5(3), 337–348. doi:10.1038/s41562-021-01056-1

Mackey, T., Purushothaman, V., Li, J., Shah, N., Nali, M., Bardier, C., & Cuomo, R. et al. (2020). Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. *JMIR Public Health and Surveillance*, 6(2), e19509–e19509. doi:10.2196/19509

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., & Adam, S. et al. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, 12(2-3), 93–118. doi:10.1080/19312458.2018.1430754

Marcec, R., Majta, M., & Likic, R. (2021). Will vaccination refusal prolong the war on SARS-CoV-2? *Postgraduate Medical Journal*, 97(1145), 143–149. doi:10.1136/postgradmedj-2020-138903

Meo, S., Bukhari, I., Akram, J., Meo, A., & Klonoff, D. (2021). COVID-19 vaccines: Comparison of biological, pharmacological characteristics and adverse effects of Pfizer/BioNTech and Moderna Vaccines. *European Review for Medical and Pharmacological Sciences*, 25(3), 1663–1669. doi:10.26355/eurrev_202102_24877

- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, 2.
- Mirzaei, A., Kazembeigi, F., Kakaei, H., Jalilian, M., Mazloomi, S., & Nourmoradi, H. (2021). Application of health belief model to predict COVID-19-preventive behaviors among a sample of Iranian adult population. *Journal of Education and Health Promotion*, 10(1), 69–69. doi:10.4103/jehp.jehp_747_20
- Mulligan, M., Lyke, K., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., & Jansen, K. et al. (2020). Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature*, 586(7830), 589–593. doi:10.1038/s41586-020-2639-4
- O., A., Gupta, S., & Hasibuddin, M. (2020). Truth Discovery in Big Data Social Media Sensing Applications. *The International Journal of Innovative Technology and Exploring Engineering*, 9(8), 999-1004. doi:10.35940/ijitee.H6311.069820
- Odlum, M., Cho, H., Broadwell, P., Davis, N., Patrao, M., Schauer, D., & Yoon, S. et al. (2020). Application of Topic Modeling to Tweets as the Foundation for Health Disparity Research for COVID-19. *Studies in Health Technology and Informatics*, 272(1), 24–27. doi:10.3233/shti200484
- Pano, T., & Kashef, R. (2020). A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. *Big Data and Cognitive Computing*, 4(4), 33.
- Pennycook, G., & Rand, D. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188(1), 39–50.
- Praveen, S., Ittamalla, R., & Deepak, G. (2021). Analyzing the attitude of Indian citizens towards COVID-19 vaccine - A text analytics study. *Diabetes & Metabolic Syndrome*, 15(2), 595–599. doi:10.1016/j.dsx.2021.02.031
- Radu, R. (2020). Fighting the ‘Infodemic’: Legal Responses to COVID-19 Disinformation. *Social Media + Society*, 6(3). doi:10.1177/2056305120948190
- Rohrer, J., Brümmer, M., Schmukle, S., Goebel, J., & Wagner, G. (2017). “What else are you worried about?” - Integrating textual responses into quantitative social science research. *PLoS One*, 12(7), e0182156–e0182156. doi:10.1371/journal.pone.0182156
- Rosnow, R. (1980). Psychology of rumor reconsidered. *Psychological Bulletin*, 87(3), 578–591. doi:10.1037/0033-2909.87.3.578
- Sakr, S., Elshawy, R., Ahmed, A., Qureshi, W., Brawner, C., Keteyian, S., & Al-Mallah, M. et al. (2017). Comparison of machine learning techniques to predict all-cause mortality using fitness data: The Henry ford exercise testing (FIT) project. *BMC Medical Informatics and Decision Making*, 17(1), 174–194. doi:10.1186/s12911-017-0566-6
- Shahnazi, H., Ahmadi-Livani, M., Pahlavanzadeh, B., Rajabi, A., Hamrah, M., & Charkazi, A. (2020). Assessing preventive health behaviors from COVID-19: A cross sectional study with health belief model in Golestan Province, Northern of Iran. *Infectious Diseases of Poverty*, 9(1), 157. doi:10.1186/s40249-020-00776-2
- Sharda, R., Delen, D., & Turban, E. (2021). *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support* (11th ed.). Pearson.
- Shen, L., Yao, R., Zhang, W., Evans, R., Cao, G., & Zhang, Z. (2021). Emotional Attitudes of Chinese Citizens on Social Distancing During the COVID-19 Outbreak: Analysis of Social Media Data. *JMIR Medical Informatics*, 9(3), e27079. doi:10.2196/27079
- Singh, M., Jakhar, A., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining*, 11(1), 33. doi:10.1007/s13278-021-00737-z
- Tai, Z., & Sun, T. (2011). The rumouring of SARS during the 2003 epidemic in China. *Sociology of Health & Illness*, 33(5), 677–693.
- Tentolouris, A., Ntanasis-Stathopoulos, I., Vlachakis, P., Tsilimigras, D., Gavriatopoulou, M., & Dimopoulos, M. (2021). COVID-19: Time to flatten the infodemic curve. *Clinical and Experimental Medicine*, 21(2), 161–165. doi:10.1007/s10238-020-00680-x

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558.

Tsao, S., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L., & Butt, Z. (2021). What social media told us in the time of COVID-19: A scoping review. *The Lancet Digital Health*, 3(3), e175–e194. doi:10.1016/S2589-7500(20)30315-0

van der Linden, S., Dixon, G., Clarke, C., & Cook, J. (2021). Inoculating against COVID-19 vaccine misinformation. *EClinicalMedicine*, 33(1), 100772. doi:10.1016/j.eclinm.2021.100772

Wang, J., Zhou, Y., Zhang, W., & Evans, R., & C., Z. (2020). Concerns Expressed by Chinese Social Media Users During the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging Data. *Journal of Medical Internet Research*, 22(11), e22152. doi:10.2196/22152

Wani, M., Agarwal, N., & Bours, P. (2021). Impact of Unreliable Content on Social Media Users during COVID-19 and Stance Detection System. *Electronics (Basel)*, 10(1), 5.

Wilson, S., & Wiysonge, C. (2020). Social media and vaccine hesitancy. *BMJ Global Health*, 5(10), e004206. doi:10.1136/bmjgh-2020-004206

Wong, M., Wong, E., Huang, J., Cheung, A., Law, K., Chong, M., & Chan, P. et al. (2021). Acceptance of the COVID-19 vaccine based on the health belief model: A population-based survey in Hong Kong. *Vaccine*, 39(7), 1148–1156. doi:10.1016/j.vaccine.2020.12.083

Yin, F., Wu, Z., Xia, X., Ji, M., Wang, Y., & Hu, Z. (2021). Unfolding the Determinants of COVID-19 Vaccine Acceptance in China. *Journal of Medical Internet Research*, 23(1), e26089. doi:10.2196/26089

Yin, H., Yang, S., & Li, J. (2020). Lecture Notes in Computer Science: Vol. 12447. *Detecting Topic and Sentiment Dynamics Due to COVID-19 Pandemic Using Social Media. Advanced Data Mining and Applications. ADMA 2020*. Springer.

Zampetakis, L., & Melas, C. (2021). The health belief model predicts vaccination intentions against COVID-19: A survey experiment approach. *Applied Psychology. Health and Well-Being*, 13(2), 1–16.

Zhou, J., Yang, S., Xiao, C., & Chen, F. (2021). Examination of Community Sentiment Dynamics due to COVID-19 Pandemic: A Case Study from a State in Australia. *SN Computer Science*, 2(3), 201–2019. doi:10.1007/s42979-021-00596-7

Mohammad Daradkeh is an Associate Professor of Business Analytics and Data Science in the Faculty of Information Technology and computer Sciences, Yarmouk University. Before joining Yarmouk University, Dr. Daradkeh has been working as a Lecturer in the Department of Informatics and Enabling Technologies at Lincoln University, New Zealand for two years. He received his PhD in Software and Information Technology from Lincoln University, New Zealand, and MSc. and BSc. in Computer Science from Yarmouk University, Jordan. His research interests lie primarily in the areas of Business Intelligence and Analytics, Visual Analytics, Decision Support Systems, and Uncertainty and Risk Management. He is currently teaching in undergraduate and graduate courses related to Decision Support systems, Business Intelligence and Analytics, Software Project Management, and Data Analytics. Dr. Daradkeh has published in Int. J. of Decision and Information Sciences, Int. J. of Business Information Systems, Journal of Organizational and End User Computing, Information Technology and People, Int. J. of Information Systems in the Service Sector, and Int. J. of Information and Knowledge Management.