BMC Complementary Medicine and Therapies

RESEARCH

Open Access

Tracking discussions of complementary, alternative, and integrative medicine in the context of the COVID-19 pandemic: a month-by-month sentiment analysis of Twitter data

Jeremy Y. Ng*, Wael Abdelkader and Cynthia Lokker

Abstract

Background: Coronavirus disease 2019 (COVID-19) is a novel infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Despite the paucity of evidence, various complementary, alternative and integrative medicines (CAIMs) have been being touted as both preventative and curative. We conducted sentiment and emotion analysis with the intent of understanding CAIM content related to COVID-19 being generated on Twitter across 9 months.

Methods: Tweets relating to CAIM and COVID-19 were extracted from the George Washington University Libraries Dataverse Coronavirus tweets dataset from March 03 to November 30, 2020. We trained and tested a machine learning classifier using a large, pre-labelled Twitter dataset, which was applied to predict the sentiment of each CAIM-related tweet, and we used a natural language processing package to identify the emotions based on the words contained in the tweets.

Results: Our dataset included 28 713 English-language Tweets. The number of CAIM-related tweets during the study period peaked in May 2020, then dropped off sharply over the subsequent three months; the fewest CAIM-related tweets were collected during August 2020 and remained low for the remainder of the collection period. Most tweets ($n = 15\,612$, 54%) were classified as positive, 31% were neutral ($n = 8803$) and 15% were classified as negative ($n = 4298$). The most frequent emotions expressed across tweets were trust, followed by fear, while surprise and disgust were the least frequent. Though volume of tweets decreased over the 9 months of the study, the expressed sentiments and emotions remained constant.

Conclusion: The results of this sentiment analysis enabled us to establish key CAIMs being discussed at the intersection of COVID-19 across a 9-month period on Twitter. Overall, the majority of our subset of tweets were positive, as were the emotions associated with the words found within them. This may be interpreted as public support for CAIM, however, further qualitative investigation is warranted. Such future directions may be used to combat misinformation and improve public health strategies surrounding the use of social media information.

*Correspondence: ngjy2@mcmaster.ca
Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

Keywords: Complementary and alternative medicine, COVID-19, Sentiment analysis, Twitter, Social media

Background

Coronavirus disease 2019 (COVID-19) is a novel infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. In December 2019 it was first discovered, having originated from Wuhan, China, and has since rapidly spread across the globe, with 220 countries reporting cases. As of March 23, 2022, over 452.2 million cases and 6.03 million deaths have been reported by the World Health Organization (WHO) [2]. Common symptoms associated with COVID-19 include fever, tiredness, and dry cough, but can also include aches and pains, nasal congestion, runny nose, sore throat or diarrhea. While some patients infected with the disease do not exhibit symptoms, COVID-19 is of great concern to global public health as approximately 5% of people who are infected will become seriously ill and need intensive care [1]. Certain health precautions such as frequent and thorough hand washing, social distancing, wearing masks, and self-isolation have been shown to reduce the spread of COVID-19 [3]. There were no proven drugs to prevent or cure COVID-19 at the outset of the pandemic [4, 5], and vaccines became available in 2021 [6, 7]. Despite this, and even with the administrative of over 10.7 billion vaccine doses administered to date, some complementary, alternative, and integrative medicines (CAIMs) have been touted as the solution [8].

According to the National Center for Complementary and Integrative Health (NCCIH), complementary and alternative medicine is defined as "health care approaches that are not typically part of conventional medical care or that may have origins outside of usual Western practice". "Complementary" refers to care in combination with conventional medicine, whereas "alternative" refers to care in place of it. "Integrative medicine" refers to bringing conventional and complementary approaches together in a coordinated way [9]. While the use of CAIMs in the context of some diseases have been shown to be effective or promising, it is also well-documented in the research literature that CAIM is sometimes promoted as a remedy for which the evidence-base is lacking [10, 11]. This is further compounded by the fact that many patients assume that CAIM is both safe and effective, even though both CAIM therapies and practitioners are generally subject to less regulation [12]. There is a growing movement of conventional and CAIM practitioners working together to support the safer and more effective uses of CAIM therapies, but concerns remain about misinformation circulated online [13–15]. Of particular interest is social media, as the body of literature that has considered its impact and growing significance as a source of health information for the general public has grown over recent years [16–18]. Emerging methodologies that have been employed to study social media content include the utilization of natural language processing (NLP), which is defined by Liddy [19] as "a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications". One of the subfields of NLP is sentiment analysis, which automatically classifies text according to the polarity (positive to negative) of the sentiments expressed therein [20]. A positive and negative sentiment can be defined as a favourable and unfavourable expression towards a subject, respectively, while a neutral sentiment represents an expression that is neither favourable nor unfavourable.

In the context of recently past pandemics, such as influenza-A (H1N1), NLP analyses of social media content (e.g. Twitter) served multiple purposes, including monitoring, predicting, and tracking levels of infection, and identifying the kinds of information circulated, distilled into content categories [21–24]. To our knowledge, a very limited amount of research has been conducted at the intersection of CAIM and social media [25, 26], while no studies have ever investigated what information surrounding CAIM is communicated across social media during any pandemics that have occurred since the inception of the Internet. In the present study, we conducted a sentiment analysis with the intent of understanding what kind of CAIM content related to COVID-19 is being generated on Twitter during the pandemic. We identified Twitter as our social media platform of choice since it is easy to use, cheap, and accessible, and the data can be easily collected in comparison to other platforms that have more restrictive privacy policies [20]. As the first study of its kind, our findings provide insight into a previously unexplored environment in the context of CAIM, that is both popular and free to patients, yet rife with quickly and continuously generated information of unassessed quality.

Methods

Approach

We used a supervised machine learning approach, in which the machine algorithm is given labelled data—a dataset that has been classified—to be used for predicting the classification of the targeted unlabelled data, in our

case CAIM-related tweets [27]. Overall, our approach consisted of the following 2 phases: 1a) training and testing a machine learning classifier using a large, prelabelled Twitter dataset, 1b) using the trained classifier to predict the sentiment class of each tweet, and 2) utilizing an NLP package to identify the emotions based on the words contained in the tweets. We first searched for CAIM-related tweets from within a set of COVID-19-filtered tweet dataset using CAIM-related search terms. All tweets analysed in this study, therefore, contained at least one CAIM-related word/term and at least one COVID-19-related word/term. We then obtained the training dataset; a large dataset of tweets that have been prelabelled based on positive and negative sentiments created by Go et al. [28] and made publicly available through the Sentiment140 website [29]. In short, a sentiment can be defined as a "positive or negative feeling", and thus training data hand-labelled by humans can be subject to a great degree of subjectivity. We chose Sentiment140 as our training dataset which mitigates this to an extent, as the tweets in the dataset were machine-labelled based on the emoticons. For example, ":)" in a tweet indicates that the tweet contains positive sentiment and ":(" indicates that the tweet contains negative sentiment. We used two supervised machine learning approaches to conduct both a sentiment analysis (using the GLMnet trained classifier [30]) and an emotion analysis (using Syuzhet NLP package in R [31]) of our CAIM-tweets dataset. Study steps are detailed in the following sections and depicted in a flowchart in Fig. 1.

Development of search strategy

Preliminary searches of Twitter-related sentiment analyses yielded no consistent or standardized method for identifying search terms. In preparation for conducting searches across a large dataset of tweets, we first searched the Twitter platform using a number of CAIM-related and COVID-19-related terms to identify the most frequently used terms. Commonly used COVID-19-related terms were relatively simple to identify, as most Twitter users used the terms "COVID", "coronavirus" or "COVID-19. Given the lack of consensus on a complete or comprehensive operational definition of CAIM [32], we browsed MeSH headings on MEDLINE and selected the most commonly used terms to refer to CAIM [33], and common CAIM systems and their respective practitioners (e.g., "homeopathy" vs. "homeopath", etc.) [9]. We excluded highly specific or specialized types of CAIM that would not typically be used by the general public (e.g., "electroacupuncture" as opposed to "acupuncture", the specific genus and species of herbs as opposed to a generic term such as "herbal medicine", etc.). A shortlist of 44 CAIM-related terms were combined with the 3 COVID-19-related terms, resulting in 132 unique Twitter searches. After applying these searches to Twitter, we looked at the recency of the use of terms to identify those most relevant to include in our final search strategy. Based on this approach, our final CAIM search strategy included the following terms: "Ayurveda", "Ayurveda medicine", "dietary supplement", "herbal", "herbal medicine", "herbal supplement", "herbal therapy", "herbalism", "herbs", "homeopathy", "homeopathic", "natural medicine", "natural medicines", "natural therapies", "natural therapy", "naturopathic medicine", "naturopathy", "traditional medicine", "traditional medicines", "vitamins", and "vitamin".

Data collection

To collect tweets at the intersection of COVID-19 and CAIM, we applied our CAIM search strategy to a COVID-19 filtered tweets dataset made available by the TweetSets website [34, 35]. TweetSets is an open-source online platform from the George Washington University (GWU) that archives Twitter datasets for research purposes. GWU Dataverse is part of the Harvard Dataverse, a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community [36]. TweetSets allows users to select, generate, and download tweet IDs from publicly available filtered tweets datasets by allowing for querying on keywords, hashtags, mentions, users, embedded media, and type of tweet (original, retweet, quotes, or reply). Through TweetSets, we accessed the Coronavirus dataset, created by Kerchner and Wrubel [37], which contained 354 903 485 COVID-19 related tweets from March 03, 2020 and November 30, 2020 as of February 03, 2021. GWU compiled the tweets by applying the keywords #Coronavirus, #Coronaoutbreak, #COVID19 using the post statuses/filter method of the Twitter stream application programming interface (API). We applied our CAIM-related search strategy to filter the Coronavirus dataset, thus identifying tweets containing both CAIM and COVID-19-related content. We limited tweets to original English-language tweets that included one or more of the CAIM-related search terms.

The TweetSets output was a condensed series of tweet IDs relating to the identity of each included tweet. To extract the text of the tweet, date of posting, user account identifiers, and tweet metadata (i.e., location coordinates, hashtags, tweets URL, retweet status, and language code), a "Hydrator" software [38] was used. This software allowed us to extract the tweet details from the tweet IDs in our search results. The output dataset was a comma-separated values (.csv) file that was imported into Microsoft Excel for data cleaning and analysis, which is described in further detail below.

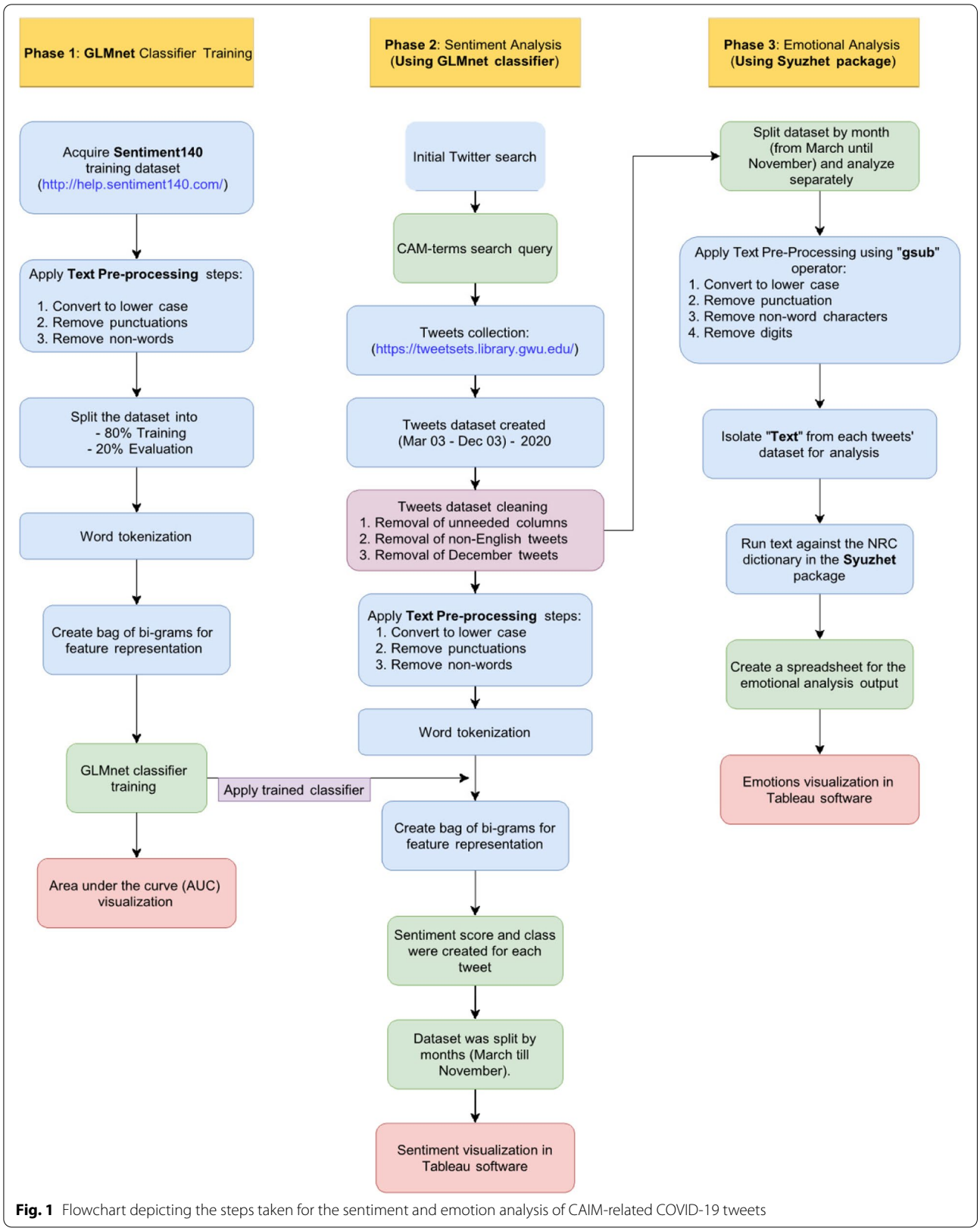

Fig. 1 Flowchart depicting the steps taken for the sentiment and emotion analysis of CAIM-related COVID-19 tweets

Sentiment analysis of CAIM-related tweets

Contextual polarity sentiment analysis involves determining the polarity of the opinion resulting in an output of positive, neutral, and negative [39]. Sentiment analyses of the collected tweets was performed in Rstudio software. The contextual polarity sentiment analysis was conducted using the Text2Vec package [40] for text processing, an R package which provides a framework for text analysis and NLP, and the GLMnet package [39] for the machine learning classifier. We used a supervised machine learning approach whereby the learning capabilities of the model was determined by a labelled training dataset. For this training, we used the Sentiment140 tweets dataset [29], which is a labelled dataset of 1.6 million twitter messages created by Go et al. [28] using machine learning to classify tweets into positive and negative based on their sentiments. The training dataset, Sentiment140, contained the targeted correct attributes (sentiment) from which the learning machine algorithm found patterns that mapped the input data attributes to the target (sentiment e.g., positivity, neutrality, negativity). The machine learning model functions by analysing the input (our tweet dataset) based on knowledge acquired from the training set, and then returning a predicted value related to the sentiment of each identified CAIM-related tweet. The training dataset was split into training and evaluation in an 80:20 ratio. Words in the training dataset were tokenized using the itoken() function in Text2Vec Package, a process of reducing a text into phrases or words called tokens. The aim of this process is to identify meaningful words in a given sentence since textual data is a stream of characters [41]. Prior to the tokenization, we applied some text pre-processing procedures to the training and testing datasets: each word was converted to lowercase, and symbols, numbers, and non-words were removed.

N-grams was used as our feature selection (i.e., the process of selecting a subset of relevant features (words, variables, attributes, or predictors)) for use in model construction. N-grams is a space reduction method that selects a subset of the dataset to identify more relevant features from the pre-processed text to improve classification quality and reduce computational complexity. N-gram is the sequence of a given number of words (N), and it is a probability model to predict the most probable word that might follow a certain sequence while preserving the word locality information; we used bi-grams which is the sequence of two words [42, 43]. For the machine to understand the text within our dataset, the text had to be vectorized in a process called text vectorization; in other words, this process transformed text into an array of numbers (vectors) to make it understandable by the machine [44]. Vectorized bi-grams were organized in a document-term matrix (DTM) —a mathematical matrix that describes the frequency of terms in a collection of texts [45]. A machine learning classifier, the algorithm for prediction of the target class label, was fit to the created DTM for training. The classifier output was set to generate fitted probabilities values for each tweet, with a score ranging between 0 and 1 (0 tending towards the most negative, 1 tending towards the most positive, and values between 0.35 and 0.65 being considered neutral [46]). We selected the regularized generalized linear model, GLMnet, as our classifier; this is an extension of the generalized linear model with built-in variable selection making them helpful in real world datasets. To decrease bias in the results of the classifier, we have used the fivefold cross validation. To evaluate the performance of our machine learning model as applied to the evaluation dataset, we determined the receiver operator characteristic (ROC) curve and area under the ROC curve (AUC).

Emotion analysis of CAIM-related tweets

To further identify the emotions relayed within our tweet dataset, we split the dataset by month (nine datasets). Analysis was performed using the Syuzhet R package, which is capable of extracting sentiment and sentiment-derived plot arcs from text using a variety of sentiment dictionaries within the package [31]. Syuzhet employs a lexicon dictionary of emotions based on the National Research Council Canada (NRC) Emotion Lexicon [47, 48]. This lexicon was created by manual annotation of a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive) accomplished by crowdsourcing. Tableau Desktop (Professional Edition) was used for the visualization of the results in terms of frequencies, percentage, and changes over time for the eight emotions.

Results

Tweet dataset

With our search terms, we identified 39 775 original tweets, of which 28 713 were posted in the English language. The most commonly used CAIM-related hashtags were #vitamin followed by #ayurveda. “Vitamin” and “vitamins” were overwhelmingly the most common CAIM-related terms followed by “herbal” and “Ayurveda”, as shown in Fig. 2. The number of CAIM-related tweets during our study period peaked in May 2020, then dropped off sharply over the subsequent three months; the fewest CAIM-related tweets were collected during August 2020 and remained low for the remainder of the collection period (Fig. 3).

Fig. 2 Word cloud depicting the most frequently mentioned words/terms contained in our subset of analyzed CAIM-related COVID-19 tweets

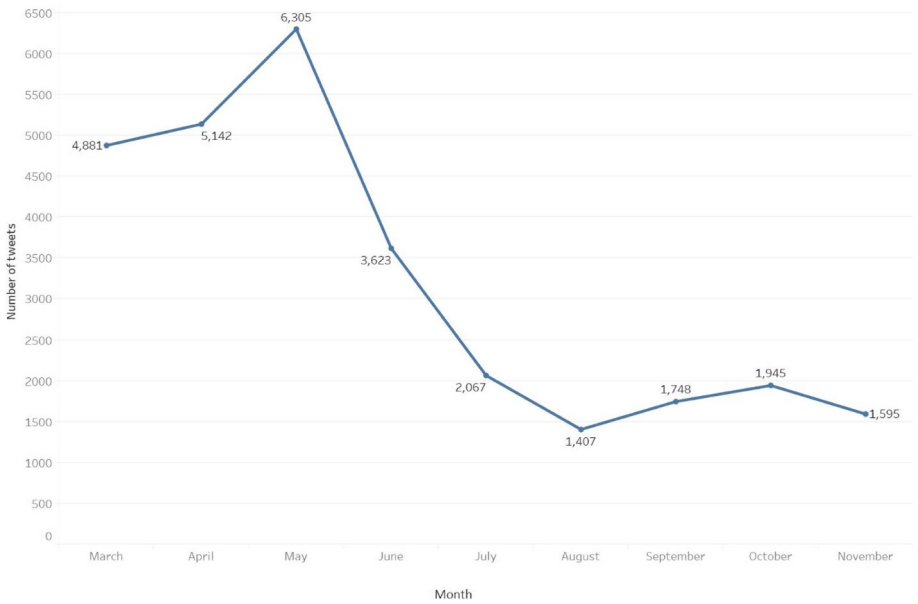

Fig. 3 Frequency of CAIM-related tweets between March 03 and November 30, 2020 shown across monthly intervals

Sentiment analysis

Our sentiment analysis algorithm using the GLM-net classifier to categorize the polarity of the tweet sentiments had an AUC of 0.894 as shown in Fig. 4A, which indicates a good ability for our classifier to distinguish between the different classes of negative and positive sentiments. Sentiments across all tweets for the 9-month period analysed were classified as positive (54.4%, $n = 15\,612$), neutral (30.7%, $n = 8803$), and negative (15%, $n = 4298$), as shown in Fig. 4B. The relative proportions of positive, negative, and neutral sentiments expressed on a month-to-month basis remained largely constant across these 9 months, as shown in Fig. 5.

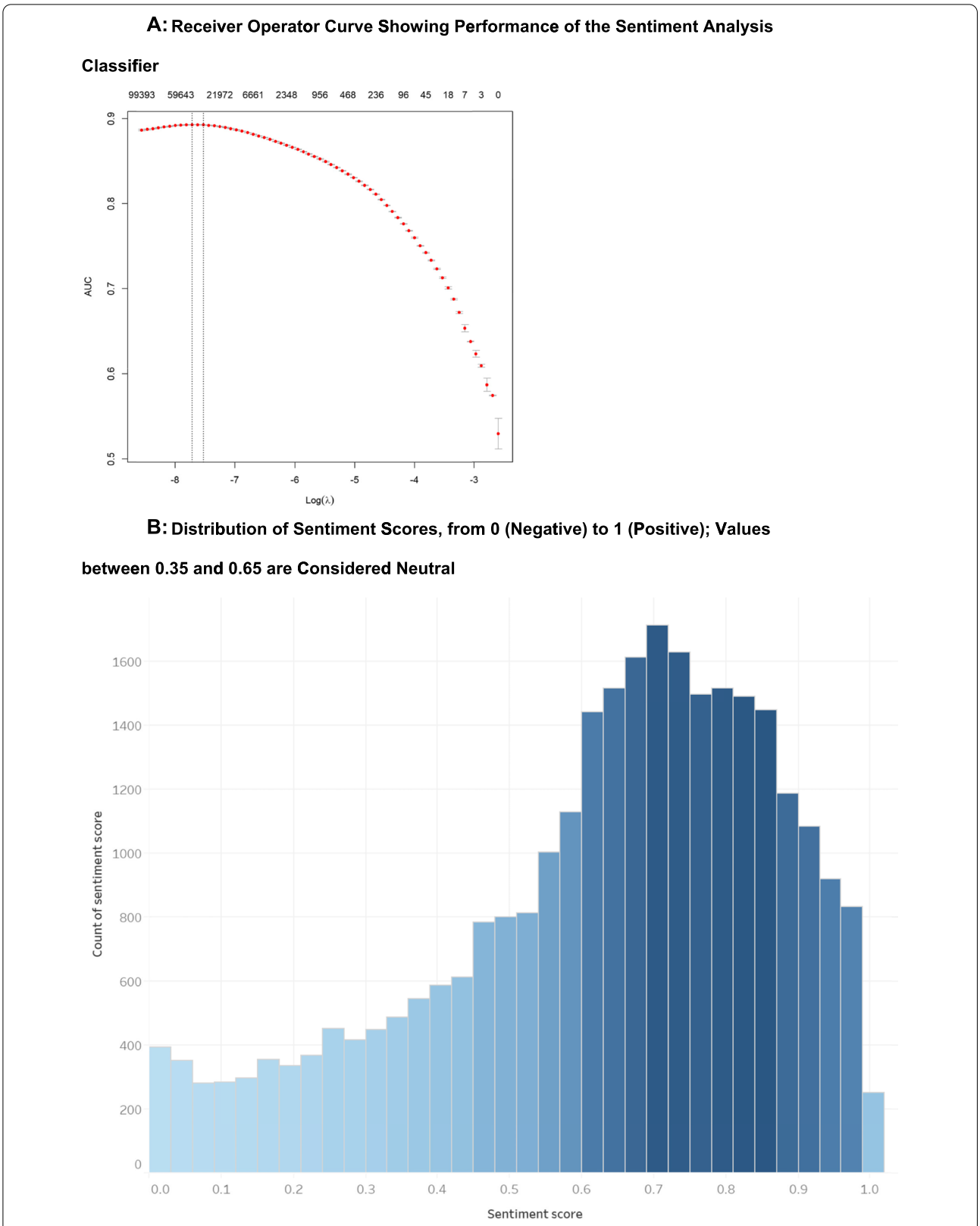

Fig. 4 **A** Receiver operator curve showing performance of the sentiment analysis classifier **B** Distribution of sentiment scores, from 0 (negative) to 1 (positive); values between 0.35 and 0.65 are considered neutral

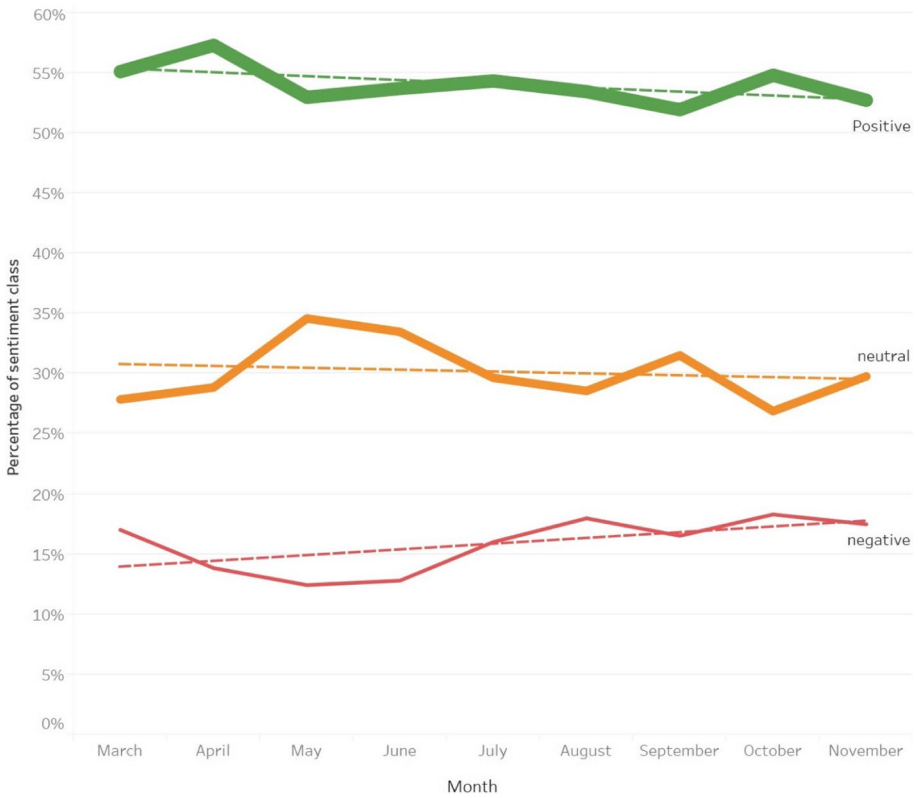

Fig. 5 Changes in sentiment from March 03 to November 30, 2020 shown across monthly intervals

Emotion analysis

When applying the algorithm employing the emotion lexicon to our tweet dataset, we were able to cross-link these emotions with text words within the tweets. The most prevalent emotion identified in the tweets was related to trust, which was associated with a total of 21,255 words. This was followed by fear ($n=16{,}410$), anticipation ($n=15{,}080$), joy ($n=11{,}407$), and sadness ($n=9669$). Anger ($n=8378$), disgust ($n=5881$), and surprise ($n=5621$) were the least represented of the eight emotions in our dataset. The relative proportions of represented emotions expressed on a month-to-month basis remained largely constant across these 9 months, as shown in Fig. 6. It is important to note that the emotions are reflective of a word itself, and not a tweet. In Table 2, we provide illustrative examples of tweets classified as positive, neutral, and negative using sentiment analysis.

Discussion

Over recent years, social media has become an increasingly popular generator and source of data that has interested a wide range of researchers [49]. The use of internet (including social media) data in studies, such as content and sentiment analyses, overcome some of the limitations of traditional social science research methods that rely on time-consuming, costly, retrospective, time-lagged, and small-scale approaches (e.g. surveys and interviews) [24, 50, 51]. In the context of pandemics, some research has even found that social media can be used to predict and detect one [52–54]. Further to this, once a pandemic has been identified, social media data can also be used to track public perceptions of the disease in question [22, 24, 55, 56]. One topic in the context of a pandemic, which has not been well-studied across social media, is the mention of CAIM. Yet, this topic is arguably of great interest given that a wide variety of CAIMs are being touted as preventative or curative against COVID-19 [57–59]. In fact, WHO Director General Tedros Adhanom Ghebreyesus at the Munich Security Conference on February 15, 2020 is quoted saying "We're not just fighting an epidemic; we're fighting an infodemic" in reference to rampant spread of misinformation, most notably across social media platforms [60].

In the present study, we conducted a sentiment and emotion analysis of Twitter data to explore what is said about CAIM in the context of COVID-19. To our knowledge, this is the first study to provide insights into the sentiments expressed by Twitter users at the intersection

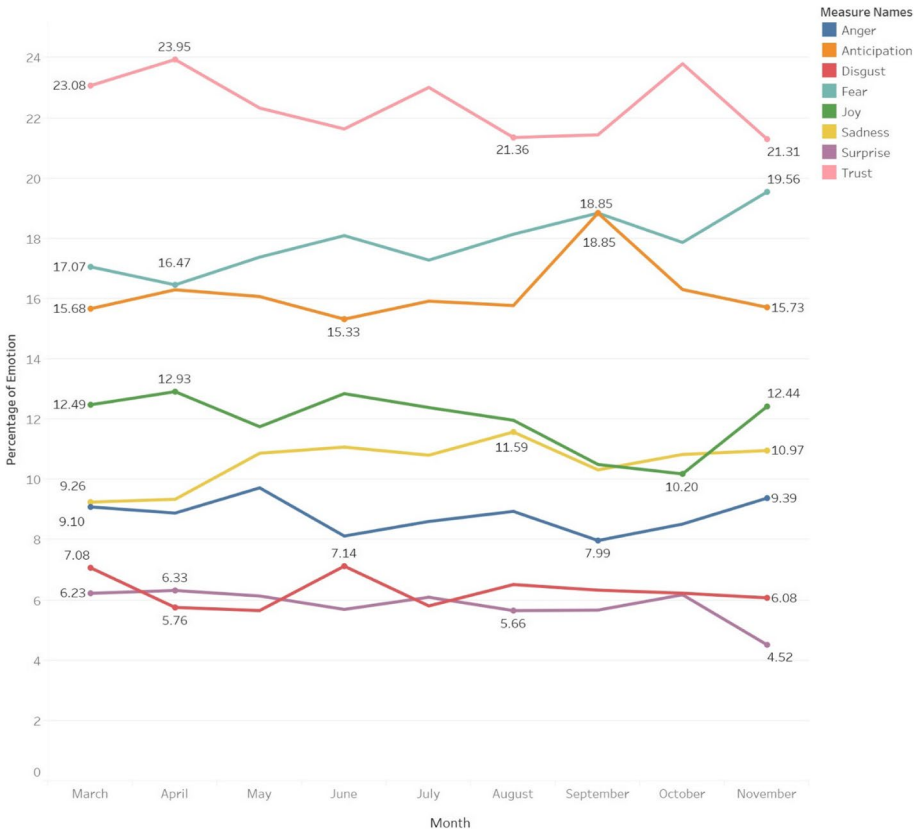

Fig. 6 Changes in emotions represented from March 03 to November 30, 2020 shown across monthly intervals

of CAIM and COVID-19. The majority of the tweets we identified and analyzed carried a generally positive sentiment. This was reflected in the emotional representation of "trust" with the highest word count in the dataset, an emotion that is frequently considered positive. We need to note the difference between the sentiment analysis of a tweet and the lexicon analysis using the Syuzhet package, as sentiment analysis is a whole tweet representation while the emotion lexicon is a word-based analysis. The latter algorithm compares words in the dataset to the NRC Sentiment and Emotional Lexicon, and it correlates words to eight standard emotions (anticipation, trust, joy, surprise, fear, sadness, anger, and disgust). From these patterns, the CAIM-related content being shared via Twitter would indicate support for CAIM interventions for COVID-19. This is in line with a plethora of published research studies that have found that the general public, across a number of different countries, tend to view CAIMs favourably and their usage continues to increase [61–65]. Over the course of our study, from March to November 2020, though the volume of tweets related to CAIM went down from the peak in May, the sentiments and emotions expressed in tweets were constant. From Table 1 and Fig. 2, as well as the illustrative tweets in Table 2, we see a focus on vitamins for prevention and treatment, which is also not entirely surprising given that across various surveys vitamins are known to be the most commonly used CAIMs [66, 67]. In fact, the 2012 National Health Interview Survey found that across all types of CAIM, natural health products (including

Table 1 Top 10 Most frequent terms from the dataset of 28 714 CAIM-related COVID-19 Tweets

Term	Tweet count
vitamin	17 527
vitamins	3736
herbal	3577
Ayurveda	3281
herbs	2100
homeopathy	1355
traditional medicine	659
homeopathic	561
herbal medicine	556
naturopathy	238

Table 2 Illustrative examples of tweets with a positive, neutral, and negative sentiments

Tweet text	Sentiment score
If yall put MORE Trust in Herbal Remedies instead of ALL this #BigPharma Prescription Shit MAYBE you could see that the only ones who will PROFIT from this #COVID19 PLANdemic IS them and the Banks! https://t.co/63G6THEgTH	positive
Even in these troubled times, do not underestimate the benefits of a simply daily #walk. Choose your location carefully but take every opportunity you can to enjoy some fresh air, sunlight and vitamin D. Learn more on Sarahs Style & Dcor #blog #COVID19 https://t.co/KFx28wcasX https://t.co/63v73iMiOH	positive
Everything you need to know about #COVID19 but your government is too afraid to tell you. Get some sunshine on your skin. Eat vitamin D rich foods, and/or supplement. Wear a mask if somewhere crowded. https://t.co/AomlgyAeTC//t.co/ https://t.co/xigTQ8SK5E https://t.co/SdKvDBcV1Q	neutral
It still surprises me that there is not more media and doctors on television telling us to strengthen our immune system, take vitamins, eat healthily, get sunlight They only seem to be talking about vaccine and drugs that are in the distant future #COVID19 #coronavirusuk	neutral
A positive test doesn't mean a healthy person is going to be sick. Also it doesn't say for the sick persons if this virus is responsible for the illness. Fear creates diseases. Be cautious but not fearful. Boost your immune system get vitamin D3 or sunlight once a day #coronavirus	negative
Disgusting NHS in go-slow on Hydroxychloroquine trials to "justify" the #Lockdown. Delays will probably needlessly kill 100's of patients. Only 2 hospitals No Zinc No Z-Pak or other antibiotic No mention of Vitamins C or D https://t.co/yJunw9PFAE #COVID19 #Covid19	negative

vitamins) were the most commonly used among Americans [68].

Comparative literature

To the authors' knowledge, informed by preliminary searches of the academic literature, the present study is the first to conduct a sentiment and emotion analysis with the intent of understanding general CAIM content related to COVID-19 generated on Twitter. If we are to look outside of this intersection of topics, however, a growing number of studies involving social media data have been published relating to COVID-19. Some of these provide a more generalized overview of public COVID-19 discussions. Xue et al. [69] used unsupervised machine learning, qualitative analysis, and sentiment analysis to understand Twitter users' discourse and psychological reactions to COVID-19, finding that while information relating to treatments and symptoms were not prevalent topics, fear of the unknown nature of the disease was dominant across all identified themes. Hung et al. [70] also applied machine learning methods to analyze data collected from Twitter including to identify the social network's dominant topics and whether the tweets expressed positive, neutral, or negative sentiments. They identified 5 main themes including: health care environment, emotional support, business economy, social change, and psychological stress. Of approximately 900 000 tweets analyzed, their sentiment analysis classified 48% of tweets as having a positive sentiment, 21% as neutral, and 31% as negative. Abd-Alrazaq et al. [71] leveraged latent Dirichlet allocation (a type of NLP) for topic modelling to identify topics discussed in the tweets relating to the COVID-19 pandemic, in addition to conducting a sentiment analysis. They identified four main themes associated with their subset of included tweets including: origin of the virus; its sources; its impact on people, countries, and the economy; and ways of mitigating the risk of infection. They also found that the mean sentiment was positive for 10 topics and negative for 2 topics (first, COVID-19-caused deaths, and second, an increase in racism). Based on their findings, they noted that a more proactive and agile public health presence on social media is warranted to combat the spread of misinformation.

Other studies have focused their objectives on identifying types or prevalence of misinformation. Mackey et al. [72] used NLP and deep learning to detect and characterize illicit COVID-19 product sales using Twitter and Instagram data. They identified a few hundred tweets and posts, respectively, containing questionable immunity-boosting treatments or involving suspect testing kits, as well as a small number of posts about pharmaceuticals that had not been approved for COVID-19 treatment. Kouzy et al. [73] conducted searches on Twitter related to COVID-19, then summarized and assessed individual tweets for misinformation in comparison to verified and peer-reviewed resources, ultimately concluding that medical misinformation and unverifiable content were being propagated at an alarming rate. In contrast, Singh et al. [74] also analysed COVID-19-related Twitter content but found that while discussions surrounding myths and links to poor quality information did exist, their presence was less dominant than other crisis-specific themes. Krawchuk et al. [75] conducted a descriptive study which detailed Twitter activity regarding spinal manipulative therapy and claims made that it increases or boosts

immunity. They found that misinformation linking spinal manipulation and increased immunity increased dramatically at the onset of the COVID-19 crisis. Lastly, Yang et al. [76] sought to understand the landscape and propagation of COVID-19 misinformation and its correction on Sina Weibo, China's largest microblogging website. While the authors did not specifically aim to capture CAIM-related information, they found that rumours surrounding false or untested therapies/measures (e.g., traditional Chinese medicine, saline water, firecrackers, and even smoking) prevention and treatment of COVID-19 were among the topics commonly circulated.

Addressing COVID-19 and CAIM misinformation on social media

Misinformation has been defined as "false and inaccurate information that is spread intentionally and unintentionally" [77] and is known to spread on social media networks easily and quickly [78, 79]. Due to the negative potential influence on people's health practices, health misinformation has received more scholarly attention, especially since the beginning of the COVID-19 pandemic [80–82]. It is particularly harmful because: 1) people are more likely to trust the information after they have been exposed to it, 2) correcting misinformation is time-consuming and resource-intensive, and 3) even after correction, it may continue to influence attitudes and behaviours, reflecting a phenomenon known as "belief echoes" [83, 84]. Correcting disinformation has become more complex and difficult as social media platforms have grown in popularity, catalysing the quick and widespread spread of misinformation. Social media networks are highly afflicted by misinformation, and it is a challenge to block or flag (re)transmission due to a lack of professional gatekeeping [85, 86]. This issue is compounded by the fact that health information seeking and scanning behaviours on social media networks increase when faced with a public health crisis [87], with COVID-19 being no exception [88, 89].

The WHO provides seven items for individuals to identify misinformation, as follows: 1) assess the source, 2) go behind the headlines, 3) identify the author, 4) check the date, 5) examine the supporting evidence, 6) check your biases, and 7) turn to fact-checkers [90]. The WHO has also published a webpage with weblinks to report misinformation found on commonly used social media platforms [91]. On a global scale, the WHO and its partners are leading three initiatives to combat misinformation online. The first involves changing social media policy and guidelines, by working with content providers such as YouTube, to reduce and remove videos containing misinformation. The second involves reporting misinformation, whereby several social media platforms granted the WHO access to fast-track reporting systems, allowing for quicker tagging and removal of content containing misinformation. Lastly, the WHO has leveraged data insights by working with YouTube, Google, and Facebook, among others, to understand where misinformation is most rampant, to target the delivery of science-based health information where it is most needed [92]. Specific to CAIM misinformation, the NCCIH offers an online resource known as "Know the Science" which provides interactive modules allowing users to learn about topics such as making sense of health research and deciding whether health news stories contain missing, misleading, or conflicting information, along with other information-related resources offered by the US National Institutes of Health and the US Centers for Disease Control and Prevention [93].

Future directions

Several future directions could be followed, based on the present study as well as emerging research in this topic area. As misinformation surrounding the COVID-19 pandemic is both rampant and pervasive on Twitter, among other social media platforms, several researchers have begun developing tools to track such misinformation. Sharma et al. [94] designed a dashboard to track misinformation on Twitter, which aims to identify false, misleading, and clickbait contents from collected tweets. Al-Rakhami et al. [95] has proposed an ensemble-learning-based framework for verifying the credibility of a vast number of tweets, which classifies tweet information based on tweet- and user-level features into two categories, either "credible" or "non-credible". Tools such as these can be applied to Twitter datasets containing information at the intersection of CAIM and COVID-19 to both compare with and validate our findings. Additionally, while our sentiment and emotion analysis provides us with insight into the polarity of sentiment and the emotions expressed in our dataset, a qualitative content analysis could identify: specific themes pertaining to this intersection of topics, trending topics, ideas most commonly linked in the text, and characterize who is generating and sharing related tweets.

Strengths and limitations

We extracted a large number of tweets that were posted over the first 9 months of the COVID-19 pandemic between March 03, 2020, and November 30, 2020 inclusive and applied two different methods to analyze the tweet dataset. We employed a supervised machine learning approach utilizing the Text2Vec package for our sentiment analysis. The purpose of this method was to acquire generalizable results built on labelled data which provided results for each tweet as a whole based on the

combination of words (respecting their locality and relation to each other), rather than a lexicon-based analysis which treats each word as a separate entity. Using the highly cited Sentiment140 dataset for training our sentiment analysis model is a strength as the dataset contains 1.6 million machine labelled tweets categorized by polarity. Finally, the Syuzhet package in R is considered a good machine learning technique to provide an emotion representation of the words within the tweets based on the NRC emotion lexicon database. We applied considerable rigour in developing our search strategy by consulting reviews of CAIM, MeSH terms, and conducting trial searches within Twitter to ensure that we identified the most relevant and used terms. It is also worth noting that few sentiment analyses published to date have analyzed or compared sentiments over multiple time periods. As opposed to capturing all tweets posted on one day or a series of days, unique to our study is the fact that we captured tweets across a period of 9 months which allowed us to compare trends over time as the pandemic progressed.

Limitations include the fact that we did not account for all CAIMs, as they represent a dynamic and wide range of therapies. This was mitigated by the preliminary searches of Twitter for the CAIMs most commonly mentioned in tweets that informed our decision on what terms to include. A further limitation is that sentiment has been classified along the continuum of positive to negative, without additional approaches to detect such linguistic elements as sarcasm, context, and complex emotions or sentiment, which are evident in the tweets illustrated in Table 2 [96]. The reliability of a model relates to its consistent performance throughout the period and conditions it is being tested [97]. Our model achieves an AUC score of 0.89 which is considered good performance for a classifier. The reliability of sentiment analysis models can be variable, with differences among them [97, 98]. We aimed to mitigate this concern by using Sentiment140 [29], a large dataset of 2.5 million labelled tweets which has been used in several other sentiment analysis studies in the context of health research [99–101]. Using this to train our model boosts confidence in its performance and reliability. During the initial phases of the study we relied on the Twitter rest/standard API, which does not allow a tweet retrieval past a certain time. Due to this limitation within the Twitter API, we relied on the Harvard Dataverse COVID-19 dataset, which had not been updated past December 03, 2020 at the time we conducted our analysis. As such, we have a narrow window of time reflected in the analyzed tweets. Given that this dataset has since been updated, in the future, we could apply our methods to discern how the sentiments and emotions in tweets have evolved as the pandemic has progressed. We limited our tweets to originals and in English. Given the global nature of the pandemic and the regional differences in CAIM treatments, we likely have missed relevant tweets. Future research on the amplification of messaging via retweets could also lead to new insights into the spread of CAIM-related content in the context of this pandemic.

Conclusions

We conducted a sentiment analysis with the objective of understanding what was being mentioned about CAIM in the context of the COVID-19 pandemic on Twitter. A total of 28 713 English-language tweets were analyzed. The most common CAIM-related hashtag used was #vitamin followed by #ayurveda. Most of the tweets were classified as positive (54%), followed by neutral (31%) and negative (15%). The most frequent emotions expressed across tweets was trust, followed by fear. Social media continues to be an important source of data that provides a range of advantages over traditional data sampling techniques, such as surveys and interviews. The use of sentiment analysis on Twitter data at the intersection of CAIM and COVID-19 provides insight into how such data is being disseminated. Our findings warrant further qualitative investigation of the emotions identified across tweets analysed, which could be used to combat the spread of misinformation and inform improved public health strategies surrounding the use of social media information.

Abbreviations

API: Application programming interface; AUC: Area under the ROC curve; CAIM: Complementary, alternative, and integrative medicine; COVID-19: Coronavirus disease 2019; DTM: Document term matrix; NCCIH: National Centre for Complementary and Integrative Health; NLP: Natural language processing; NRC: National Research Council Canada; ROC: Receiver operator characteristic; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; WHO: World Health Organization.

Acknowledgements

JYN was awarded a Research Scholarship and an Entrance Scholarship from the Department of Health Research Methods, Evidence and Impact, Faculty of Health Sciences at McMaster University.

Authors' contributions

JYN: co-designed and conceptualized the study, collected and analysed data, co-drafted the manuscript, and gave final approval of the version to be published. WA: collected and analysed data, co-drafted the manuscript, and gave final approval of the version to be published. CL: co-designed and conceptualized the study, collected and analysed data, co-drafted the manuscript, and gave final approval of the version to be published.

Funding

This study was unfunded.

Availability of data and materials

All relevant data are included in this manuscript.

Declarations

Ethics approval and consent to participate

This study involved a sentiment analysis of publicly available data on Twitter. This study was deemed by the Hamilton Integrated Research Ethics Board to be exempt from their oversight. All methods were carried out in accordance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2018).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 1 April 2021 Accepted: 7 April 2022
Published online: 13 April 2022

References

1. World Health Organization. Coronavirus disease (COVID-19) [Internet]; 2020. [cited 2022 Mar 18]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19
2. World Health Organization. Coronavirus disease (COVID-19) pandemic [Internet]; 2020. [cited 2022 Mar 18]. Available from: https://covid19.who.int/
3. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet. 2020. https://doi.org/10.1016/j.jvs.2020.07.040.
4. WHO Solidarity Trial Consortium. Repurposed antiviral drugs for COVID-19—interim WHO SOLIDARITY trial results. N Engl J Med. 2020. https://doi.org/10.1056/NEJMoa2023184.
5. Simonovich VA, Burgos Pratx LD, Scibona P, Beruto MV, Vallone MG, Vázquez C, Savoy N, Giunta DH, Pérez LG, Sánchez MD, Gamarnik AV. A randomized trial of convalescent plasma in COVID-19 severe pneumonia. N Engl J Med. 2020. https://doi.org/10.1056/NEJMoa2031304.
6. Lurie N, Saville M, Hatchett R, Halton J. Developing COVID-19 vaccines at pandemic speed. N Engl J Med. 2020;382(21):1969–73. https://doi.org/10.1056/NEJMp2005630.
7. Peiris M, Leung GM. What can we expect from first-generation COVID-19 vaccines? The Lancet. 2020;396(10261):1467–9. https://doi.org/10.1016/S0140-6736(20)31976-0.
8. National Center for Complementary and Integrative Health. In the news: coronavirus and "alternative" treatments [Internet]; 2020. [cited 2020 Dec 10]. Available from: https://www.nccih.nih.gov/health/in-the-news-coronavirus-and-alternative-treatments
9. National Center for Complementary and Integrative Health. Complementary, alternative, or integrative health: what's in a name? [Internet]; 2018. [cited 2020 Dec 10]. Available from: https://www.nccih.nih.gov/health/complementary-alternative-or-integrative-health-whats-in-a-name
10. Bianchini C, Truccolo I, Bidoli E, Group CI, Mazzocut M. Avoiding misleading information: a study of complementary medicine online information for cancer patients. Libr Inf Sci Res. 2019;41(1):67–77. https://doi.org/10.1016/j.lisr.2019.02.006.
11. Murdoch B, Zarzeczny A, Caulfield T. Exploiting science? A systematic analysis of complementary and alternative medicine clinic websites' marketing of stem cell therapies. BMJ Open. 2018;8(2): e019414. https://doi.org/10.1136/bmjopen-2017-019414.
12. Cancer.org. Why are complementary and alternative therapies harder to evaluate? [Internet]; 2020. [cited 2020 Dec 10]. Available from: https://www.cancer.org/treatment/treatments-and-side-effects/complementary-and-alternative-medicine/complementary-and-alternative-methods-and-cancer/why-cam-is-hard-to-evaluate.html
13. Chen AT, Taylor-Swanson L, Buie RW, Park A, Conway M. Characterizing websites that provide information about complementary and integrative health: Systematic search and evaluation of five domains. Interactive Journal of Medical Research. 2018;7(2): e9803. https://doi.org/10.2196/ijmr.9803.
14. Sharma V, Holmes JH, Sarkar IN. Identifying complementary and alternative medicine usage information from internet resources. Methods Inf Med. 2016;55(04):322–32. https://doi.org/10.3414/ME15-01-0154.
15. Mazzocut M, Truccolo I, Antonini M, Rinaldi F, Omero P, Ferrarin E, et al. Web conversations about complementary and alternative medicines and cancer: Content and sentiment analysis. J Med Internet Res. 2016;18(6): e120. https://doi.org/10.2196/jmir.5521.
16. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract. 2013;14(1):15–23. https://doi.org/10.1177/1524839911405850.
17. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res. 2009;11(4): e48. https://doi.org/10.2196/jmir.1249.
18. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res. 2013;15(4): e85. https://doi.org/10.2196/jmir.1933.
19. Liddy ED. Natural language processing. Encyclopaedia of library and information science [Internet]. 2nd ed. New York: Marcel Decker, Inc.; 2001. [cited 2020 Dec 10]. Available from: https://surface.syr.edu/cnlp/11/
20. Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: Review of the methods used. JMIR Public Health Surveillance. 2018;4(2): e43. https://doi.org/10.2196/publichealth.5789.
21. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PloS One. 2011;6(5). https://doi.org/10.1371/journal.pone.0019467
22. Szomszor M, Kostkova P, St Louis C. Twitter informatics: Tracking and understanding public reaction during the 2009 swine flu pandemic. 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology; 2011 Aug 22; Vol. 1, p. 320–3; IEEE. https://doi.org/10.1109/WI-IAT.2011.311
23. Kostkova P, Szomszor M, St. Louis C. #swineflu: The use of twitter as an early warning and risk communication tool in the 2009 swine flu pandemic. ACM Transactions on Management Information Systems (TMIS). 2014 Jul 1;5(2):1–25. https://doi.org/10.1145/2597892
24. Chew C, Eysenbach G. Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PloS One. 2010;5(11). https://doi.org/10.1371/journal.pone.0014118
25. Delgado-López PD, Corrales-García EM. Influence of internet and social media in the promotion of alternative oncology, cancer quackery, and the predatory publishing phenomenon. Cureus. 2018 May;10(5). https://doi.org/10.7759/cureus.2617
26. Marcon AR, Klostermann P, Caulfield T. Chiropractic and spinal manipulation therapy on Twitter: Case study examining the presence of critiques and debates. JMIR Public Health Surveill. 2016;2(2): e5739. https://doi.org/10.2196/publichealth.5739.
27. Talabis MRM, McPherson R, Miyamoto I, Martin JL, Kaye D. Chapter 1: Analytics defined. In: Information security analytics. Elsevier; 2015. p. 1–12. https://doi.org/10.1016/B978-0-12-800207-0.00001-0
28. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12), 2009. https://www-cs-faculty.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf
29. Sentiment140?. What is Sentiment140? [Internet]. [cited 2020 Dec 10]. Available from: http://help.sentiment140.com/
30. Glmnet Vignette. [Internet]. [cited 2020 Dec 10]. Available from: https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html
31. Jockers M. Package 'syuzhet' [Internet]; 2017. [cited 2020 Dec 10]. Available from: https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html
32. Cochrane Complementary Medicine. Operational definition of complementary medicine [Internet]. [cited 2020 Dec 10]. Available from: https://cam.cochrane.org/operational-definition-complementary-medicine
33. Ng JY, Boon HS, Thompson AK, Whitehead CR. Making sense of "alternative", "complementary", "unconventional" and "integrative" medicine:

Exploring the terms and meanings through a textual analysis. BMC Complement Altern Med. 2016;16(1):1–8. https://doi.org/10.1186/s12906-016-1111-3.
34. TweetSets. Twitter datasets for research and archiving [Internet]. [cited 2020 Dec 10]. Available from: https://tweetsets.library.gwu.edu/
35. Justin Littman. TweetSets. Zenodo [Internet]; 2018. https://doi.org/10.5281/zenodo.1289426
36. Harvard Dataverse Website. [Internet]; n.d. [cited 2020 Dec 10]. Available from: https://dataverse.harvard.edu/
37. Kerchner D, Wrubel L. Coronavirus Tweet IDs [Internet]; 2020. Harvard Dataverse, V7. https://doi.org/10.7910/DVN/LW0BTB
38. Documenting the Now. Hydrator [Computer Software] [Internet]; 2020. [cited 2020 Dec 10]. Available from: https://github.com/docnow/hydrator
39. Muhammad A, Wiratunga N, Lothian R. Contextual sentiment analysis for social media genres. Knowl-Based Syst. 2016;15(108):92–101. https://doi.org/10.1016/j.knosys.2016.05.032.
40. Selivanov D, Wang Q. text2vec: Modern text mining framework for R. Computer software manual (R package version 0.4. 0) [Internet]; 2016. [cited 2020 Dec 10]. Available from: https://CRAN.R-project.org/package=text2vec.
41. Verma T, Renu R, Gaur D. Tokenization and filtering process in RapidMiner. International Journal of Applied Information Systems. 2014;7(2):16–8. https://doi.org/10.5120/ijais14-451139.
42. Houvardas J, Stamatatos E. N-gram feature selection for authorship identification. In: Euzenat J., Domingue J, editors. Artificial intelligence: Methodology, systems, and applications. AIMSA 2006. Lecture Notes in Computer Science, vol 4183. Berlin, Heidelberg: Springer; 2006. https://doi.org/10.1007/11861461_10
43. Järvelin A, Järvelin A, Järvelin K. s-grams: Defining generalized n-grams for information retrieval. Inf Process Manage. 2007;43(4):1005–19. https://doi.org/10.1016/j.ipm.2006.09.016.
44. Sueno HT, Gerardo BD, Medina RP. Converting text to numerical representation using modified Bayesian vectorization technique for multi-class classification. International Journal. 2020 Jul;9(4). https://doi.org/10.30534/ijatcse/2020/211942020
45. Madsen RE, Sigurdsson S, Hansen LK, Larsen J. Pruning the vocabulary for better context recognition. 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541); n.d., 2, p. 1439–44. https://doi.org/10.1109/IJCNN.2004.1380163
46. Nimirthi P, Krishna PV, Obaidat MS, Saritha V. A framework for sentiment analysis based recommender system for agriculture using deep learning approach. In: Social network forensics, cyber security, and machine learning; 2019. p. 59–66. Singapore: Springer. https://doi.org/10.1007/978-981-13-1456-8_5
47. National Research Council Canada. The sentiment and emotion lexicons [Internet]. [cited 2020 Dec 10]. Available from: http://sentiment.nrc.ca/lexicons-for-research/
48. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. Comput Intell. 2013;29(3):436–65. https://doi.org/10.1111/j.1467-8640.2012.00460.x.
49. Kapoor KK, Tamilmani K, Rana NP, Patil P, Dwivedi YK, Nerur S. Advances in social media research: Past, present and future. Inf Syst Front. 2018;20(3):531–58. https://doi.org/10.1007/s10796-017-9810-y.
50. Mavragani A. Infodemiology and infoveillance: Scoping review. J Med Internet Res. 2020;22(4): e16206. https://doi.org/10.2196/16206.
51. Eysenbach G. Infodemiology and infoveillance: Tracking online health information and cyberbehavior for public health. Am J Prev Med. 2011;40(5):S154–8. https://doi.org/10.1016/j.amepre.2011.02.006.
52. Samaras L, García-Barriocanal E, Sicilia MA. Comparing social media and Google to detect and predict severe epidemics. Sci Rep. 2020;10(1):1–1. https://doi.org/10.1038/s41598-020-61686-9.
53. Al-Garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: A systematic review. J Biomed Inform. 2016;1(62):1–1. https://doi.org/10.1016/j.jbi.2016.05.005.
54. Ritterman J, Osborne M, Klein E. Using prediction markets and Twitter to predict a swine flu pandemic. 1st International Workshop on Mining Social Media; 2009 Nov 9; Vol. 9, p. 9–17. https://www.pure.ed.ac.uk/ws/portalfiles/portal/13611111/Ritterman_Osborne_Klein_2009_Using_Prediction_Markets_and_Twitter_to_Predict_a_Swine_Flu_Pandemic.pdf
55. Jain VK, Kumar S. An effective approach to track levels of influenza-A (H1N1) pandemic in India using Twitter. Procedia Computer Science. 2015;1(70):801–7. https://doi.org/10.1016/j.procs.2015.10.120.
56. Lampos V, Cristianini N. Tracking the flu pandemic by monitoring the social web. 2010 2nd International Workshop on Cognitive Information Processing; 2010 Jun 14; p. 411–16; IEEE. https://doi.org/10.1109/CIP.2010.5604088
57. Caulfield T. Pseudoscience and COVID-19-we've had enough already. Nature. 2020. https://doi.org/10.1038/d41586-020-01266-z.
58. Desta TT, Mulugeta T. Living with COVID-19-triggered pseudoscience and conspiracies. Int J Public Health. 2020;65(6):713–4. https://doi.org/10.1007/s00038-020-01412-4.
59. Larson HJ. Blocking information on COVID-19 can fuel the spread of misinformation. Nature. 2020:306-. https://doi.org/10.1038/d41586-020-00920-w
60. Zarocostas J. How to fight an infodemic. The Lancet. 2020;395(10225):676. https://doi.org/10.1016/S0140-6736(20)30461-X.
61. Klein SD, Torchetti L, Frei-Erb M, Wolf U. Usage of complementary medicine in Switzerland: Results of the Swiss health survey 2012 and development since 2007. PLoS ONE. 2015;10(10): e0141985. https://doi.org/10.1371/journal.pone.0141985.
62. National Center for Complementary and Integrative Health. Statistics on complementary and integrative health approaches. [Internet]; 2020. [cited 2020 Dec 10]. Available from: https://www.nccih.nih.gov/research/statistics-on-complementary-and-integrative-health-approaches
63. Esmail N. Complementary and alternative medicine: Use and public attitudes 1997, 2006, and 2016. Vancouver: Fraser Institute, 2017. [Internet]; 2020. [cited 2020 Dec 10]. Available from: https://www.fraserinstitute.org/sites/default/files/complementary-and-alternative-medicine-2017.pdf
64. Lobera J, Rogero-García J. Scientific appearance and homeopathy. Determinants of trust in complementary and alternative medicine. Health Communication. 2021;36(10):1278–85. https://doi.org/10.1080/10410236.2020.1750764.
65. Islahudin F, Shahdan IA, Mohamad-Samuri S. Association between belief and attitude toward preference of complementary alternative medicine use. Patient Preference and Adherence. 2017;11:913.
66. O'Brien SK, Malacova E, Sherriff JL, Black LJ. The prevalence and predictors of dietary supplement use in the Australian population. Nutrients. 2017;9(10):1154. https://doi.org/10.3390/nu9101154.
67. Moore J, McClain A, Hong MY. Dietary supplement use in the United States: Prevalence, trends, pros, and cons. Nutr Today. 2020;55(4):174–81. https://doi.org/10.1097/NT.0000000000000402.
68. Clarke TC, Black LI, Stussman BJ, Barnes PM, Nahin RL. Trends in the use of complementary health approaches among adults: United States, 2002–2012. National Health Statistics Reports. 2015 Feb 10(79):1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4573565/
69. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: using latent Dirichlet allocation for topic modeling on Twitter. PLoS ONE. 2020;15(9): e0239441. https://doi.org/10.1371/journal.pone.0239441.
70. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, Hon SD, Park J, Dang P, Lipsky MS. Social network analysis of COVID-19 sentiments: Application of artificial intelligence. J Med Internet Res. 2020;22(8): e22590. https://doi.org/10.2196/22590.
71. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res. 2020;22(4): e19016. https://doi.org/10.2196/19016.
72. Mackey TK, Li J, Purushothaman V, Nali M, Shah N, Bardier C, Cai M, Liang B. Big data, natural language processing, and deep learning to detect and characterize illicit COVID-19 product sales: Infoveillance study on Twitter and Instagram. JMIR Public Health Surveill. 2020;6(3): e20794. https://doi.org/10.2196/20794.
73. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, Zarka J, Traboulsi C, Akl EW, Baddour K. Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on Twitter. Cureus. 2020 Mar 13;12(3). https://doi.org/10.7759/cureus.7255
74. Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, Padden C, Vanarsdall R, Vraga E, Wang Y. A first look at COVID-19 information and

misinformation sharing on Twitter. arXiv preprint arXiv:2003.13907. 2020 Mar 31. https://doi.org/10.48550/arXiv.2003.13907
75. Kawchuk G, Hartvigsen J, Harsted S, Nim CG, Nyirö L. Misinformation about spinal manipulation and boosting immunity: An analysis of Twitter activity during the COVID-19 crisis. Chiropractic & Manual Therapies. 2020;28(1):1–3. https://doi.org/10.1186/s12998-020-00319-4.
76. Yang Q, Luo Z, Li M, Liu J. Understanding the landscape and propagation of COVID-19 misinformation and its correction on Sina Weibo. Global Health Promotion. 2021 Sep 11:17579759211035053.
77. Chen L, Wang X, Peng TQ. Nature and diffusion of gynecologic cancer–related misinformation on social media: analysis of tweets. J Med Internet Res. 2018;20(10): e11515. https://doi.org/10.2196/11515.
78. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med. 2019;1(240): 112552. https://doi.org/10.1016/j.socscimed.2019.112552.
79. Sylvia Chou WY, Gaysynsky A, Cappella JN. Where we go from here: health misinformation on social media. Am J Public Health. 2020;110(S3):S273–5. https://doi.org/10.2105/AJPH.2020.305905.
80. Lewandowsky S, Ecker UK, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: Continued influence and successful debiasing. Psychological Science in the Public Interest. 2012;13(3):106–31.
81. Enders AM, Uscinski JE, Klofstad C, Stoler J. The different forms of COVID-19 misinformation and their consequences. The Harvard Kennedy School Misinformation Review. 2020. https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37366466
82. Bavel JJ, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, Crockett MJ, Crum AJ, Douglas KM, Druckman JN, Drury J. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav. 2020;4(5):460–71. https://doi.org/10.1038/s41562-020-0884-z.
83. Thorson E. Belief echoes: The persistent effects of corrected misinformation. Polit Commun. 2016;33(3):460–80. https://doi.org/10.1080/10584609.2015.1102187.
84. Sharevski F, Alsaadi R, Jachim P, Pieroni E. Misinformation warnings: Twitter's soft moderation effects on covid-19 vaccine belief echoes. Comput Secur. 2022;1(114): 102577. https://doi.org/10.1016/j.cose.2021.102577.
85. Gallo JA, Cho CY. Social Media: Misinformation and Content Moderation Issues for Congress. Congressional Research Service Report. 2021;46662. https://crsreports.congress.gov/product/pdf/R/R46662
86. Gimpel H, Heger S, Olenberger C, Utz L. The effectiveness of social norms in fighting fake news on social media. J Manag Inf Syst. 2021;38(1):196–221. https://doi.org/10.1080/07421222.2021.1870389.
87. Pew Research Center. Social media and health [Internet]. 2010. [cited 2022 Mar 18]. Available from: https://www.pewresearch.org/internet/2010/03/24/social-media-and-health/
88. Statista. Social media use during COVID-19 worldwide - statistics & facts curve [Internet]; 2022. [cited 2022 Mar 18]. Available from: https://www.statista.com/topics/7863/social-media-use-during-coronavirus-covid-19-worldwide/#dossierKeyfigures
89. Singh S, Dixit A, Joshi G. Is compulsive social media use amid COVID-19 pandemic addictive behavior or coping mechanism? Asian Journal of Psychiatry. 2020;54:10229.
90. World Health Organization. Let's flatten the infodemic curve [Internet]; 2022. [cited 2022 Mar 18]. Available from: https://www.who.int/news-room/spotlight/let-s-flatten-the-infodemic-curve
91. World Health Organization. How to report misinformation online [Internet]; 2022. [cited 2022 Mar 18]. Available from: https://www.who.int/campaigns/connecting-the-world-to-combat-coronavirus/how-to-report-misinformation-online
92. World Health Organization. Combatting misinformation online [Internet]; 2022. [cited 2022 Mar 18]. Available from: https://www.who.int/teams/digital-health-and-innovation/digital-channels/combatting-misinformation-online
93. National Center for Complementary and Integrative Health. Know the science [Internet]; 2022. [cited 2022 Mar 18]. Available from: https://www.nccih.nih.gov/health/know-science
94. Sharma K, Seo S, Meng C, Rambhatla S, Liu Y. Covid-19 on social media: Analyzing misinformation in twitter conversations. arXiv preprint arXiv:2003.12309. 2020 Mar 26. https://doi.org/10.48550/arXiv.2003.12309
95. Al-Rakhami MS, Al-Amri AM. Lies kill, facts save: Detecting COVID-19 misinformation in Twitter. IEEE Access. 2020;26(8):155961–70. https://doi.org/10.1109/ACCESS.2020.3019600.
96. Sarsam SM, Al-Samarraie H, Alzahrani AI, Wright B. Sarcasm detection using machine learning algorithms in Twitter: A systematic review. International Journal of Market Research. 2020;62(5):578–98.
97. Lawrence L. Reliability of sentiment mining tools: A comparison of semantria and social mention (Bachelor's thesis, University of Twente). 2014. http://essay.utwente.nl/65302/1/Lawrence_BA_MB.pdf
98. Jussila J, Vuori V, Okkonen J, Helander N. Reliability and perceived value of sentiment analysis for Twitter data. In: Strategic Innovative Marketing 2017 (pp. 43–48). Springer, Cham. https://doi.org/10.1007/978-3-319-56288-9_7
99. Korkontzelos I, Nikfarjam A, Shardlow M, Sarker A, Ananiadou S, Gonzalez GH. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from Tweets and forum posts. Journal of Biomedical Informatics. 2016;62:148–58. https://doi.org/10.15265/IY-2017-029.
100. Garcia K, Berton L. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Appl Soft Comput. 2021;1(101): 107057. https://doi.org/10.1016/j.asoc.2020.107057.
101. Imran AS, Daudpota SM, Kastrati Z, Batra R. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related Tweets. IEEE Access. 2020;28(8):181074–90. https://doi.org/10.1109/ACCESS.2020.3027350.

Publisher's Note