

## Early detection of fraudulent COVID-19 products from Twitter chatter

### Author List:

\*Abeed Sarker, PhD<sup>1</sup>  
Sahithi Lakamana, MS<sup>1</sup>  
Ruqi Liao, BS<sup>2</sup>  
Aamir Abbas, MS, MD<sup>3</sup>  
Yuan-Chi Yang, PhD<sup>1</sup>  
Mohammed Ali Al-Garadi, PhD<sup>1</sup>

### Corresponding Author:

Abeed Sarker, PhD  
101 Woodruff Circle  
Office 4101  
Atlanta, GA 30322  
[abeed@dbmi.emory.edu](mailto:abeed@dbmi.emory.edu);  
Phone: +1(602) 474 6203

<sup>1</sup> Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, United States

<sup>2</sup> H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, United States

<sup>3</sup> Heinz College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, PA, United States

### Keywords (MeSH)

coronavirus (D017934)  
COVID-19 drug treatment (C000705127)  
social media (D061108)  
public health surveillance (D062486)

### Word count:

**Abstract:** 236 words; **Body:** 2578 words

## ABSTRACT

Social media have served as lucrative platforms for misinformation and for promoting fraudulent products for the treatment, testing and prevention of COVID-19. This has resulted in the issuance of many warning letters by the United States Food and Drug Administration (FDA). While social media continue to serve as the primary platform for the promotion of such fraudulent products, they also present the opportunity to identify these products early by employing effective social media mining methods. In this study, we employ natural language processing and time series anomaly detection methods for automatically detecting fraudulent COVID-19 products early from Twitter. Our approach is based on the intuition that increases in the popularity of fraudulent products lead to corresponding anomalous increases in the volume of chatter regarding them. We utilized an anomaly detection method on streaming COVID-19-related Twitter data to detect potentially anomalous increases in mentions of fraudulent products. Our unsupervised approach detected 34/44 (77.3%) signals about fraudulent products earlier than the FDA letter issuance dates, and an additional 6/44 (13.6%) within a week following the corresponding FDA letters. Our proposed method is simple, effective and easy to deploy, and do not require high performance computing machinery unlike deep neural network-based methods.

## 1. INTRODUCTION

As of 7<sup>th</sup> September, 2021, over 220 million confirmed COVID-19 cases have been reported globally, with over 41 million reported cases in the United States (US) alone [1]. As governments and public health agencies around the globe enacted efforts to mitigate the impact of the pandemic, one persistent problem has been the opportunistic promotion of fraudulent products claiming to treat, prevent, test and/or cure COVID-19 infections. There have been numerous reports of adverse health events caused by toxic exposures to fraudulent products that have no scientific evidence supporting their use [2], [3]. In response to the emergence of many fraudulent products, the US Food and Drug Administration (FDA) has issued warning letters [4]. These warning letters are typically issued after the products become popular and many people have already been exposed to them. Since it is not possible to advertise fraudulent products on television or via reliable news sources, social media platforms have been exploited for the mass promotion of such products. In fact, promotional content regarding such products over social networks, such as Twitter, is only a subset of the misinformation spread through these platforms, which has been referred to as an *infodemic* [5], [6]. The fraudulent products are often promoted directly via the social media accounts (eg., Twitter, Facebook) of the entities profiting from their sales, and, if the promotions gain traction, information about them are circulated by other social media users. Consequently, information regarding the products spread through social networks in analogous patterns as other types of misinformation, including those about COVID-19 [7]. There is thus the need to develop toxicovigilance tools that can automatically identify potentially fraudulent COVID-19 products early and generate alerts. While social networks provide fertile grounds for the proliferation of misinformation about fraudulent products, they also provide opportunities for responding to diverse challenges posed by the pandemic, and one potential utility of social media is the automated real-time surveillance of fraudulent COVID-19 products.

In this paper, we demonstrate that chatter about fraudulent products on Twitter, if curated systematically via natural language processing (NLP) and data-centric methods, can provide detectable early signals. We utilize publicly available streaming data from the Twitter COVID-19 application programming interface (API), which was specifically created by the company to aid COVID-19 related research [8]. Specifically, using Twitter data, we show that social media based surveillance can detect many fraudulent products early, relative to the FDA warning issuance dates. Our approach to detecting fraudulent products is based on a simple intuition—that products that gain popularity among Twitter users, following their successful promotion, will exhibit increases in their mentions in COVID-19 related chatter. These abrupt increases in the frequency of mentions are likely to be detectable by time series anomaly detection methods. It is also likely that products that gain *relatively* higher popularity will exhibit anomalous increases of *relatively* higher magnitudes in their mentions among all COVID-19 related Twitter chatter. We describe our methods and results in the following sections.

## 2. METHODS

### 2.1 Data collection

We collected data using the COVID-19 streaming API of Twitter [8]. This API was made available by Twitter specifically for supporting COVID-19 related research, and it does not impose throughput limitations or daily/monthly quotas. Consequently, we were able to collect all tweets that mentioned the COVID-19 related keywords and phrases (eg., *coronavirus*, *covid19* and *covid*) [8]. Streaming data was stored in real-time in a *mongodb* database hosted on the Google Cloud

Platform. The collection of data was continuous with only minor down times that were necessary for system modifications or updates.

## 2.2 Product detection

The list of products and entities were manually collected from the FDA website [4]. The products included were advertised as treatments/cures, tests or preventative measures for COVID-19. We curated a comprehensive list of entity names, products, FDA letter dates, person(s) who owned the entities or the products, websites and social media profiles (if any). We curated this information for a total of 183 letters issued by the FDA. Each warning letter was manually reviewed. From these, we manually curated a set of product names and/or entity names that were potentially used for promotion over social media. If the same product was mentioned in multiple letters, we only included the first mention of the product or entity and the corresponding date, excluding the later ones. We also manually curated key words and phrases that were likely to be used to refer to the products or entities on Twitter. The full list of products and entities and their earliest letter dates are provided in Table 1.

**Table 1.** Key phrases included in this study along with their types and the date of the first letter mentioning each.

Number	Key phrase	Type	First detected letter date
1	Antimicrobial solution	Treatment	11/02/2020
2	Aromatherapy	Treatment	03/6/2020
3	Bee products	Treatment	10/23/2020
4	Berberine	Treatment	10/23/2020
5	Betterfly	Treatment	09/01/2020
6	Bioflavonoids	Treatment	10/23/2020
7	Biomagnetism	Treatment	08/19/2020
8	Chlorine dioxide		04/08/2020
9	Cod liver oil	Treatment	05/25/2020
10	Colostrum	Treatment	05/26/2020
11	Covid-19 rapid test kit	Test kit	06/10/2020
12	Iodine products	Treatment	06/10/2020
13	Curativa	Treatment	06/25/2020
14	Elderberry syrup	Treatment	11/10/2020
15	Elderberry tincture	Treatment	03/06/2020
16	Hypochlorous acid	Treatment	11/02/2020
17	Kratom	Treatment	05/15/2020
18	Niacin product	Treatment	09/01/2020
19	Magnetic therapy	Treatment	08/19/2020
20	Methylene blue	Treatment	06/29/2020
21	Nad+	Treatment	05/06/2020

22	Nephron pharmaceuticals	Treatment	09/22/2020
23	Novabay	Entity	11/02/2020
24	Oracare	Treatment	11/18/2020
25	Pro breath	Treatment	11/18/2020
26	Quercetine	Treatment	6/15/2020
27	Santiste	Entity	04/27/2020
28	Superblue silver immune gargle	Treatment	04/09/2020
29	Traditional Chinese medicine	Treatment	05/08/2020
30	Transdermal patch/defend™ patch	Treatment	04/27/2020
31	Umbilical cord blood	Treatment	06/04/2020
32	Vapore	Treatment	07/30/2020
33	Vidacord	Treatment	06/04/2020
34	Xosomes	Treatment	06/04/2020
35	Ayurvedic products	Treatment	04/13/2020
36	Colloidal silver	Treatment	03/06/2020
37	Corona-cure	Treatment	03/26/2020
38	Essential oil	Treatment	03/06/2020
39	Eupatorium perfoliatum	Treatment	03/06/2020
40	Grapefruit seed extract	Treatment	05/26/2020
41	Salt therapy	Treatment	03/30/2020
42	Supersilver whitening toothpaste	Treatment	04/09/2020
43	Super C	Treatment	04/21/2020
44	Vivify	Entity	03/06/2020

Since product and entity names are often misspelled by social media subscribers, we generated potential spelling variants or misspellings of the products and entities using a data-centric tool [9]. The variant generation tool uses a combination of semantic and lexical similarity measures to automatically identify common misspellings and spelling variants of terms/phrases, including multi-word expressions. Our past work revealed that such lexical expansion strategies are capable of significantly increasing retrieval/detection rates from Twitter [10]. Examples of product names extracted from the warning letters and their automatically-generated lexical variants are shown in Table 2. We included all products/entities and their spelling variants that had at least 10 mentions in our collected data. We excluded key phrases that were mentioned less than 10 times because such low occurrences indicated that the corresponding products/entities were either not promoted over Twitter or never actually gained popularity on the platform. We counted the number of mentions of each product/entity, including their spelling variants, from the entire collected dataset. Counts of spelling variants were grouped with the original products/entities. Daily counts were normalized by the total number of posts collected on the same days. The daily relative frequencies were represented as the number of mentions per 1000 tweets.

**Table 2.** Fraudulent product names extracted from the FDA warning letters and their automatically-generated lexical variants.

Product	Spelling variants
chlorine dioxide	chlorinedioxide  chloride dioxide  chorine dioxide  clorine dioxide  clorinedioxide
fortify humic beverage concentrate	fortify humic beverage concentrates  fortify humic beverage cocentrate
electrify fulvic beverage concentrate	electrify fulvic beverage cocentrate  electrify fulvic beverage concetrate  electrify fulvic beverage concentrates
supersilver whitening toothpaste	supersilver whitening toothpast  supersilver whitening toothpastes  supersilver whitening tooth paste
superblue fluoride free toothpaste	superblue fluoride free tooth paste  superblue fluoride free toothpastes  superblue fluoride free toothpast
prefense hand sanitizers	prefense handsanitizers  prefense hand sanitizes  prefense hand sanitiers  prefense hand andsanitizers  prefense hand  prefense hand handsantizers  prefense hand handsanitizers  prefense handsantizer  prefense handsanitizers  prefense  prefense hand santitizers  prefense handsanitisers  prefense handsanitizer
covid-19 cough syrup	covid 19 cough syrups  covid 19 coughsyrup  covid 19 cough syrup  covid 19 cough coughsyrup
ncov19 spike protein	ncov19 spike spike protein  ncov19 spike spikeproteins  ncov19 spike protei  ncov19 spikey proteins  ncov19 spike spikeprotein  ncov19 spikeprotien  ncov19 spike proteins  ncov19 spike spikey proteins  ncov19 spikeprotein  ncov19 spikeproteins  ncov19 spike spikeprotien

### 2.3 Detecting anomalies

We applied a 14-day moving average filter to construct a smooth line representing the daily mention frequencies, and anomalies or outliers were detected relative to this moving average line. For each day, the *residual* for standard deviation calculation was computed by subtracting the 14-day moving average from the relative frequency per 1000 tweets on that day. For a given day ( $n$ ), the standard deviation for the day ( $\sigma_n$ ), is computed progressively, given as:

$$\sigma_n = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu_i)}{n - 1}}$$

where  $x_i$  is the relative frequency for day  $i$  and  $\mu_i$  is the 14-day moving average on day  $i$ . Thus, the standard deviation computed for a given day includes all the data points starting from day 1. The standard deviation for the first day (February 19, 2020) for any product is by definition 0. This may potentially give the anomaly detection approach an unfair advantage by increasing the sensitivity of detection in the early days easier. Therefore, we artificially added a non-zero standard

deviation on day 1, computed as  $\frac{std((\frac{x_1+1}{x_1} + \sum \mu_{1..4}))}{k}$ , where  $x_1$  is the product mention frequency on day 1,  $X_1$  is the total number of tweets collected on day 1,  $std()$  is the standard deviation function, and  $\mu_{1..4}$ , are the moving averages over the first 4 days. The value is divided by  $k$  to adjust for the  $k \times std()$  function that is applied to compute the boundaries beyond which

a data point would be considered an outlier ( $k=3$  in our experiments). This artificial initial bias that we added therefore decreases the chances of our approach to detect outliers early on in the timelines, and makes the task of detecting anomalies slightly harder, particularly for products that have relatively low number of mentions. For some products, for example, there are many days with 0 mentions early on in their timelines, but the added bias causes the progressive standard deviation to be non-zero. For 3 products with letter issue dates in March 2020, this added bias caused the method to miss early outliers that are detectable without adding the bias. Specific details are provided in the supplementary material (S3). Minimum value for daily relative frequency was set at 0.001 (*i.e.*,  $k \times 0.001$  served as the minimum threshold for outlier detection).

The chosen window size (14) and standard deviation (3), for which we report results in this paper, were relatively conservative choices for signal detection. We also performed experiments with multiple window sizes (7, 10, and 14) and standard deviation thresholds (2, 2.5, and 3) to study how the anomaly detection performance varied based on these parameters. Slight variations in window sizes and standard deviations did not impact overall performance.

#### 2.4 Evaluation

Data points that had a distance of more than 3 standard deviations from the moving average were considered to be outliers (*i.e.*, signals). For each key phrase, the date of the first outlier was compared with the FDA letter issuance date to determine if the signal was detected earlier, within 1 week, or later than the FDA letter issuance date. System percentage accuracy was computed using the formula:  $\frac{\#early}{\#total}$ . For products that were mentioned in multiple letters, our approach was only considered to be successful in early detection if the outlier was detected prior to the first mention date. Thus, the reported system performance is actually likely to be lower than in practice.

### 3. RESULTS

The issue dates of the letters ranged from March 6, 2020 to June 22, 2021. Through manual review of each letter, we identified 221 potential keywords/phrases that were either associated with the products (*e.g.*, product names) or the entities selling them. From this set, we excluded key phrases collected after the year 2020. Some products were promoted by different entities at different times, causing them to be repeated in the warning letters. Since our primary objective was to assess the possibility of early detection, we excluded repeated key phrases, keeping only their first occurrences ( $n=56$ ). Furthermore, since our focus was to detect products that gained popularity via promotion on Twitter, we excluded key phrases that were mentioned less than 10 times including their lexical variants ( $n=12$ ). 44 key phrases met all the inclusion criteria. Table 1 presents all 44 keywords, their types (*i.e.*, product or entity), and the FDA letter issuance dates. The full curated data along with additional information is available as supplementary material (Table S1).

We included a total of 577,872,350 COVID-19 related tweets in our analysis, which were collected from February 19, 2020 to December 31, 2020. We computed the daily counts of the key phrases (along with their spelling variants, if any). Increases in key phrase mention counts that were higher than 3 standard deviations from the 14-day moving average of mentions were flagged as potential ‘signals’. 43 out of the 44 key phrases showed anomalous increases in their mentions at some point of time within our collected data. For 34 out of the 44 key phrases (77.3%), signals of anomalous increases in chatter were detectable prior to the FDA letter issuance dates. An additional 6 (13.6%) key phrases had anomalous increases within seven days of the FDA letter issuance dates. Daily counts for the products, their 3-standard deviation ranges, and the moving



averages are shown in Figure 1 of supplementary material. The figure specifically shows the daily normalized frequencies for each product from 02/19/2020 to 12/31/2020. 14-day moving average is represented by a blue line and the range of three standard deviations is shown as a shaded area. A vertical line in each figure (red) shows the FDA letter issuance date and orange-black circles represent the data points falling outside the three standard deviation range. Products detected earlier than the FDA letter issuance dates are indicated by a lime-green background and the rest are indicated by a beige background. The ordering of the products is the same as the ordering in Table 1. The daily counts for all 44 key phrases are provided as supplementary material (S2; `daily_counts.xlsx`).

## 4. DISCUSSION

### 4.1 Related work

Our work is not the first to explore the utility of social media as a potential source for detecting fraudulent COVID-19 products. In recent works, unsupervised NLP methods such as topic modeling and supervised methods such as text classification have been proposed for the automatic detection of such products from social media data [11]–[13]. Others focused more broadly on detecting misinformation using social media or Internet-based data [14], [15]. However, these studies did not take into account the time factor. Typically, once the FDA issues warning about a fraudulent product, there is a rise in chatter regarding the product, but such rises are driven by media coverage or increased public awareness. We observed this phenomenon for most products included in the study, particularly the ones detected within one week of the FDA letter issuance dates. To the best of our knowledge, our approach is the first to attempt to detect fraudulent treatments early. The proposed approach is also simple, computationally inexpensive as it relies on fundamental characteristics of social media chatter (*ie.*, increases in volume of chatter about a particular topic resulting from increases in its popularity), and unsupervised (*ie.*, no training data required).

### 4.2 Limitations

There are several potential limitations of the proposed approach. First, it requires data that is not rate limited (*eg.*, data from the standard Twitter streaming API). Anomalous increases may not be detectable from rate limited streams since large increases in volume are likely to be dampened by the APIs. For real-time fraudulent product candidate detection, deployment needs to be on streaming data, although it is also possible to periodically run the anomaly detection scripts on stored, static data. Second, we were only able to calculate the percentage of early detection within our given sample, and based on the current data, we were unable to realistically estimate confidence intervals for the percentage values reported. Third, the anomaly detection approach relies on characteristic abrupt increases in chatter volumes about a given topic. It is possible that some fraudulent products may gain popularity gradually, causing the normalized counts to never go beyond the standard deviation threshold. In such cases, varying the window size (*eg.*, using 7-day moving averages) and/or lowering the standard deviation thresholds may improve the detection capability of the method. However, lowering the standard deviation threshold is also likely to result in larger numbers of false positives—an aspect that we did not take into account in this study. We believe that not taking false positives into account in the current study is justifiable since in practical settings, all signals associated with noun phrases would be reviewed by experts, and so it is perhaps better if the method is biased in favor of *recall* (*ie.*, more true and false positives) rather than *precision*.



We also do not address candidate fraudulent substance detection in this study. Several mechanisms can be used for detecting candidates including but not limited to named entity recognition (likely to be high precision but low recall), simple part-of-speech tagging to identify noun phrases (high recall, low precision), and topic modeling methods that identify possible topics from texts (low recall, high precision). We intend to explore these strategies in future work. Even without this component, we believe our approach is an improvement over past studies that did not take into account the warning letter dates. Finally, since the daily counts are normalized by the total number of tweets on the same day, it is possible that large increases in absolute counts of specific key phrases are not detectable due to equal or larger increases in the total volume of posts on the same day.

## **5. CONCLUSION**

The emergence of fraudulent products associated with COVID-19 has been a significant problem in the fight against the pandemic. Social media has served as platforms for advertising and promoting fraudulent products. While social media makes it easier for opportunist entities to promote and sell fraudulent products, this resource may also be used to conduct surveillance of fraudulent substances. In this paper, we showed that it is possible to detect many fraudulent products potentially early from Twitter data. Our simple approach employed a time series anomaly detection method for detecting anomalous increases in mentions of fraudulent substances in Twitter chatter, and obtained promising performance. Future work will focus on deploying the NLP pipeline and improve upon the limitations described in the Discussion section.

## FUNDING

N/A

## COMPETING INTERESTS STATEMENT

None declared.

## CONTRIBUTIONS

AS conceived the study, designed the experiments and prepared the manuscript. SL conducted data collection, experiments and image generation. RL and AA prepared and reviewed the dataset. YY and MA to review the research and help with preparation of the manuscript.

## REFERENCES

- [1] Centers for Disease Control and Prevention, “CDC COVID Data Tracker,” *Trends in Number of COVID-19 Cases and Deaths in the US Reported to CDC, by State/Territory*, 2020. [https://covid.cdc.gov/covid-data-tracker/#trends\\_totalandratedeathsseven-dayrate](https://covid.cdc.gov/covid-data-tracker/#trends_totalandratedeathsseven-dayrate) (accessed Feb. 19, 2021).
- [2] L. Turner, “Preying on Public Fears and Anxieties in a Pandemic: Businesses Selling Unproven and Unlicensed ‘Stem Cell Treatments’ for COVID-19,” *Cell Stem Cell*, vol. 26, no. 6, pp. 806–810, Jun. 2020, doi: 10.1016/J.STEM.2020.05.003.
- [3] H. Reihani, M. Ghassemi, M. Mazer-Amirshahi, B. Aljohani, and A. Pourmand, “Non-evidenced based treatment: An unintended cause of morbidity and mortality related to COVID-19,” *Am. J. Emerg. Med.*, vol. 39, p. 221, Jan. 2021, doi: 10.1016/J.AJEM.2020.05.001.
- [4] Food and Drug Administration, “Fraudulent Coronavirus Disease 2019 (COVID-19) Products | FDA,” 2020. <https://www.fda.gov/consumers/health-fraud-scams/fraudulent-coronavirus-disease-2019-covid-19-products> (accessed Aug. 09, 2021).
- [5] T. L. I. Diseases, “The COVID-19 infodemic,” *Lancet. Infect. Dis.*, vol. 20, no. 8, p. 875, Aug. 2020, doi: 10.1016/S1473-3099(20)30565-X.
- [6] M. Cinelli *et al.*, “The COVID-19 social media infodemic,” *Sci. Rep.*, vol. 10, no. 1, p. 16598, Dec. 2020, doi: 10.1038/s41598-020-73510-5.
- [7] H. J. Larson, “The biggest pandemic risk? Viral misinformation,” *Nature*, vol. 562, no. 7727, p. 309, Oct. 2018, doi: 10.1038/D41586-018-07034-4.
- [8] Twitter, “Twitter Developer Platform | COVID-19 stream,” *COVID-19 Stream*, 2020. <https://developer.twitter.com/en/docs/labs/covid19-stream/overview> (accessed Aug. 08, 2021).
- [9] A. Sarker, “LexExp: a system for automatically expanding concept lexicons for noisy biomedical texts,” *Bioinformatics*, vol. 37, no. 16, pp. 2499–2501, Aug. 2021, doi: 10.1093/BIOINFORMATICS/BTAA995.
- [10] A. Sarker and G. Gonzalez-Hernandez, “An unsupervised and customizable misspelling generator for mining noisy health-related text sources,” *J. Biomed. Inform.*, 2018.
- [11] T. Ken *et al.*, “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 Heal. Surveill* 2020;6(3)e20794 <https://publichealth.jmir.org/2020/3/e20794>, vol. 6, no. 3, p. e20794, Aug. 2020, doi: 10.2196/20794.
- [12] A. Fittler, L. Adeniye, Z. Katz, and R. Bella, “Effect of Infodemic Regarding the Illegal Sale of Medications on the Internet: Evaluation of Demand and Online Availability of Ivermectin during the COVID-19 Pandemic,” *Int. J. Environ. Res. Public Heal.* 2021, Vol. 18, Page 7475, vol. 18, no. 14, p. 7475, Jul. 2021, doi: 10.3390/IJERPH18147475.
- [13] K. Madhusudhana, “TEXT CLASSIFICATION MODELS FOR AUTOMATIC DETECTION OF FAKE COVID PRODUCTS AND NEWS ON SOCIAL MEDIA by,” University of Buffalo, Buffalo, New York, 2021.
- [14] T. K. Mackey, V. Purushothaman, M. Haupt, M. C. Nali, and J. Li, “Application of unsupervised machine learning to identify and characterise hydroxychloroquine misinformation on Twitter,” *Lancet Digit. Heal.*, vol. 3, no. 2, pp. e72–e75, Feb. 2021, doi: 10.1016/S2589-7500(20)30318-6.
- [15] A. Bin Shams *et al.*, “Web Search Engine Misinformation Notifier Extension (SEMiNExt): A Machine Learning Based Approach during COVID-19 Pandemic,” *Healthc.* 2021, Vol. 9, Page 156, vol. 9, no. 2, p. 156, Feb. 2021, doi: 10.3390/HEALTHCARE9020156.



