



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Whether the weather will help us weather the COVID-19 pandemic: Using machine learning to measure twitter users' perceptions

Marichi Gupta^{a,b}, Aditya Bansal^{a,c}, Bhav Jain^{a,d}, Jillian Rochelle^{a,e}, Atharv Oak^{a,d},
 Mohammad S. Jalali^{a,f,*}

^a MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA, USA

^b The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

^c Indian Institute of Technology Delhi, New Delhi, Delhi, India

^d Massachusetts Institute of Technology, Cambridge, MA, USA

^e Northwestern University, Evanston, IL, USA

^f Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Keywords:

Individuals' perceptions
 Opinion mining
 Topic modeling
 SARS-CoV-2 transmission
 Machine learning

ABSTRACT

Objective: The potential ability for weather to affect SARS-CoV-2 transmission has been an area of controversial discussion during the COVID-19 pandemic. Individuals' perceptions of the impact of weather can inform their adherence to public health guidelines; however, there is no measure of their perceptions. We quantified Twitter users' perceptions of the effect of weather and analyzed how they evolved with respect to real-world events and time.

Materials and Methods: We collected 166,005 English tweets posted between January 23 and June 22, 2020 and employed machine learning/natural language processing techniques to filter for relevant tweets, classify them by the type of effect they claimed, and identify topics of discussion.

Results: We identified 28,555 relevant tweets and estimate that 40.4 % indicate uncertainty about weather's impact, 33.5 % indicate no effect, and 26.1 % indicate some effect. We tracked changes in these proportions over time. Topic modeling revealed major latent areas of discussion.

Discussion: There is no consensus among the public for weather's potential impact. Earlier months were characterized by tweets that were uncertain of weather's effect or claimed no effect; later, the portion of tweets claiming some effect of weather increased. Tweets claiming no effect of weather comprised the largest class by June. Major topics of discussion included comparisons to influenza's seasonality, President Trump's comments on weather's effect, and social distancing.

Conclusion: We exhibit a research approach that is effective in measuring population perceptions and identifying misconceptions, which can inform public health communications.

1. Introduction

1.1. Background and significance

Since the beginning of the outbreak, one of the major questions has been whether the transmission of SARS-CoV-2 is seasonal, such as with influenza, [1] MERS [2], or SARS [3]. While there was limited research and consensus at the beginning of the pandemic on the impact of weather and seasonality on the transmission of SARS-CoV-2 [4–12], a growing body of evidence has suggested that the effect of weather conditions is modest and that weather alone is not sufficient to quench

the pandemic [13]. Despite (limited) academic consensus, what the public thinks is unknown, which motivated our research.

As COVID-19 has disrupted the global population, many have turned to social media platforms such as Twitter to navigate COVID-19. While Twitter's effectiveness at disseminating information can be leveraged to share public health information for social good, it can also promote misinformation [14]. As the virus continues to spread, chatter online has increased in volume, and one particularly contentious topic of discussion surrounds the myth that heat can effectively kill the virus [15]. While it is not uncommon for public opinion to contradict scientific literature, the continuous debate, uncertainty, and lack of consensus

* Corresponding author at: Harvard Medical School, 101 Merrimac St, Room 1032, Boston, MA, 02114, United States.

E-mail address: msjalali@mgh.harvard.edu (M.S. Jalali).

<https://doi.org/10.1016/j.ijmedinf.2020.104340>

Received 24 September 2020; Received in revised form 3 November 2020; Accepted 9 November 2020

Available online 10 November 2020

1386-5056/© 2020 Elsevier B.V. All rights reserved.

among experts exacerbated this specific public misconception [16,17]. As public knowledge of pandemic guidelines can influence the adoption of recommended behaviors [14], measuring and analyzing the social perception of the weather's impact on COVID-19 may help predict adherence to public health policy and guidelines. Machine learning and natural language processing techniques have historically proved effective for opinion mining on Twitter [18,19], which motivated our use of them.

1.2. Objectives

This study examined Twitter users' perceptions concerning the weather's effect on the spread of COVID-19 with natural language processing and machine learning techniques. Specifically, the research objectives were to identify: (1) the perceived impact of weather in relevant tweets and classify them accordingly, and (2) if and how these perceptions changed throughout the pandemic. To investigate these, we trained a support vector machine classifier to measure what proportion of tweets claim there is an effect of weather, and exhibit time-series trends for a subset of relevant tweets. To detect perceptions outside of this effect-oriented framework, we employed unsupervised learning to discover unexpected discussion topics. Our purpose then is to understand how English-language Twitter users believe the weather will impact COVID-19 and identify any misconceptions held by such users.

This study is one of many to use machine learning and natural language processing to retrieve information about public perception through social media for public health purposes, [18,19] but the first to study the perception of the weather's impact on COVID-19. We hope that this work can inform public policy and research as the COVID-19 pandemic response continues.

2. Materials and methods

2.1. Tweet collection

Using Twitter's Premium application programming interface (API) for historical search, we collected 166,005 tweets from January 23 to June 22, 2020 with the query "(coronavirus OR covid OR covid19) AND weather." This query checked all tweet components for a match, including the tweet's text, the text of any attached articles or media, and any URL text included with the tweet. We only collected English-language quoted or original tweets, not retweets. We did not limit the data to any specific location. Whenever possible, we deduce the location a tweet is posted from by checking the location of the tweet author or the tagged location of the tweet itself (see Supplementary S2 for more details). For tweets replying to or quoting another tweet, we fetched the text of the other tweet. For tweets sharing an article, we collected the article headline and description as displayed on Twitter. The tweet text, article data, and any replied-to/quoted tweets were then merged for analysis. Fig. 1 presents our research method and the flow of its processes, which are discussed below.

2.2. Reducing corpus to relevant tweets

2.2.1. Rule-based filtering

Initially, we cleaned tweets by removing any non-alphanumeric characters (including emojis), mentions of other users, and hashtags at the end of the tweet, and then we further standardized with lemmatization and stemming (see S2). Following common techniques used for social media analysis in other domains, [20] we employed rule-based filtering to narrow our corpus down and remove noise. The rule-based filtering consisted of three rules applied sequentially. First, we filtered out false positives coming from the sheer popularity of our keywords (e.g., a tweet commenting on pleasant weather and ending with "#coronavirus") and removed tweets where the keywords were split across different parts of the tweet (e.g., "weather" only appearing in the article

text, and "covid" only in the tweet itself). Second, we discarded tweets using "weather" as a verb or idiomatically (e.g., "under the weather"). Finally, we restricted the tweets to those posted by individuals, not news organizations, since individual perception was the focus of study. The strengths of these three rules were verified manually (see S3).

2.2.2. Relevancy classification

Overall, rule-based filtering reduced the corpus from 166,005 to 84,201 tweets. We used machine learning to further reduce the corpus to tweets that had insightful relationships between weather and the spread of COVID-19.

2.2.2.1. Annotation. To create training data for the classifier, two annotators (JR, BJ) labeled a set of tweets based on pre-defined inclusion criteria, which defined a tweet as relevant if it referenced a causal or correlative relation between weather and coronavirus spread, and irrelevant otherwise. Tweets presenting a causal relationship declared the weather to have a direct impact on the spread of COVID-19 (e.g., high temperatures killing the virus) while a correlative relationship declared an indirect impact (e.g., reduced social distancing during pleasant weather). Irrelevant tweets mentioned weather and COVID-19 but did not establish a relationship between them (e.g., extreme weather causing additional strain in hard-hit areas). Annotators marked a shared pilot set of 100 tweets to calibrate on these criteria (see S4.1). After resolving any discrepancies, annotators labeled a full set of training data for our machine learning classifiers. Of the 84,201 tweets remaining after the rule-based filtering, a random sample of 2768 tweets (which included the pilot set) was annotated and used for training the Relevancy Classifier.

2.2.2.2. Natural language processing and featurization. Text featurization was used to convert tweets into meaningful vectors for machine learning analysis. Three vectorization techniques were used: Bag of Words (BOW), Term Frequency-Inverse Document Frequency (TF-IDF), and Embeddings from Language Models (ELMo), a state-of-the-art technique that utilizes word embeddings [21]. ELMo factors in the surrounding context for each word (i.e., the words around it) for its vectorization, while BOW and TF-IDF do not [22]. For BOW and TF-IDF, we removed stop words (set of commonly used words which do not contribute to the context of the tweet) and also words that only appeared in 1% of all tweets or less.

We tested 11 classification models for performance on relevancy classification: Ridge Classifier, Logistic Regression, k-Nearest Neighbors, Support Vector Machine, Logistic Regression with Gradient Descent, Support Vector Machine with Gradient Descent, Multinomial Naïve Bayes, Complement Naïve Bayes, Bernoulli Naïve Bayes, Random Forest Classifier, and Decision Trees (see S5). We used Scikit-learn's machine learning libraries [23].

We performed a five-fold outer cross-validation on our training dataset to select the optimal model with five-fold inner cross-validation to find the ideal hyperparameters (see S4). For each of our models, we evaluated and reported the Area under the Precision-Recall curve (AUC-PR) and Area under the Receiver Operating Characteristic curve (AUC-ROC)—for definitions, see [24]. Both metrics are presented, but we chose to optimize with respect to AUC-PR since it provides a better assessment of model performance for imbalanced datasets, where AUC-ROC be overly optimistic [24–26]. We took the best performing model to be our "Relevancy Classifier" that produced the corpus for analysis, both for the claimed effect of weather and for topics of discussion.

3. Analyzing tweets for effect

Using the rule-based filtering methods and the Relevancy Classifier described above, we filter the full set of 166,005 tweets to a set of 28,555

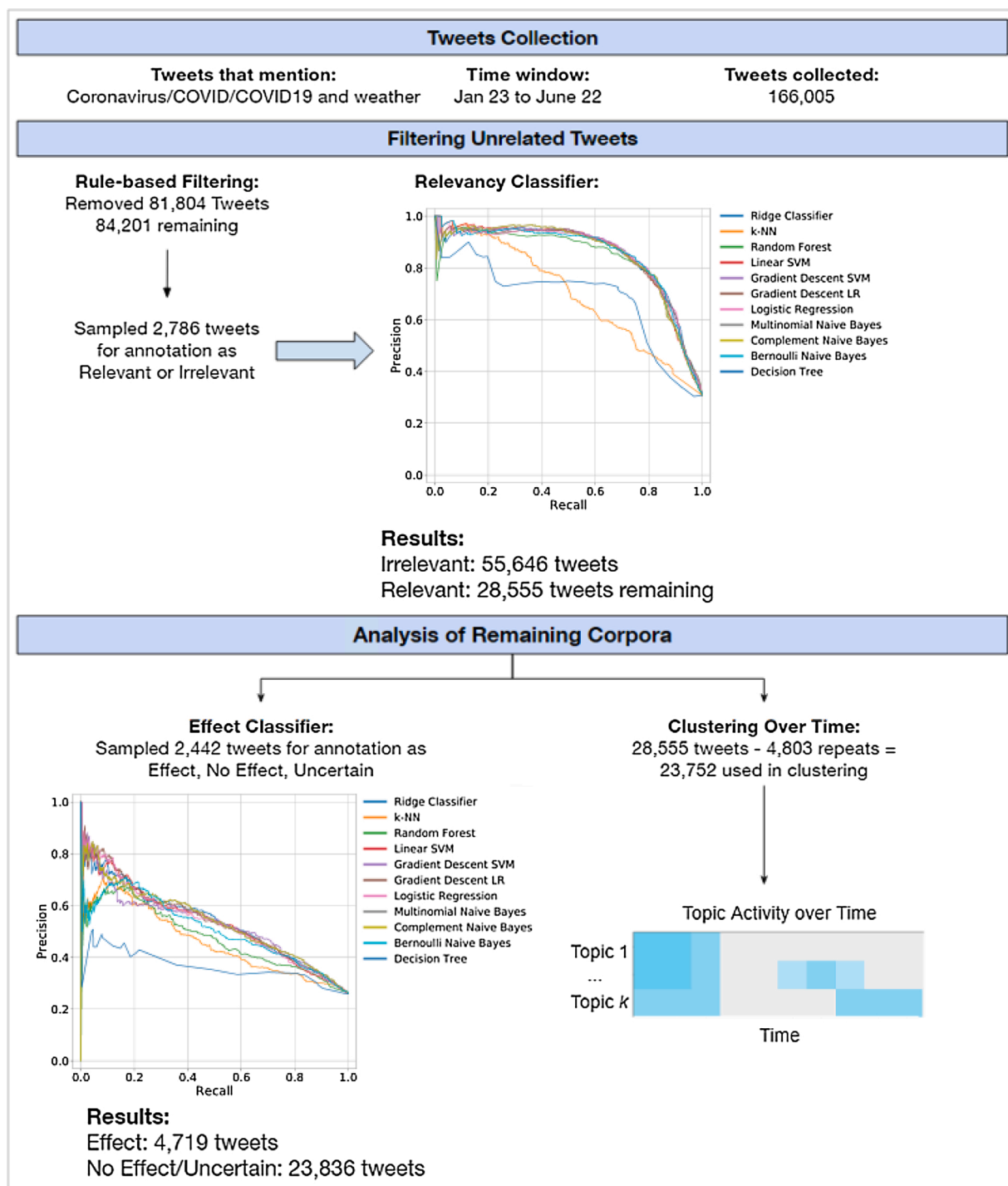


Fig. 1. Flow diagram of filtering and machine learning processes.

tweets, which are used for both the effect and topic modeling analyses (see Fig. 1). To classify tweets based on the type of effect the user expected the weather to have on the spread of COVID-19, we trained another machine learning classifier.

3.1. Effect classification

3.1.1. Annotation

We first annotated a new batch of tweets (distinct from the relevancy annotation set) based on if they claimed weather to have some effect and used this as training data. After calibrating on a pilot set of 200 tweets

(see S4.2), annotators (JR, BJ, MG) first labeled tweets into one of three categories: “effect,” where the tweet suggested that weather had an impact on COVID-19; “no effect,” where the tweet suggested weather had no impact; and “uncertain,” where the tweet was uncertain to the effect or made no clear claim to an effect.

Additionally, within the “effect” category, tweets were labeled based on whether the tweet suggested COVID-19 would: i) improve with warmer weather, ii) worsen with warmer weather, iii) improve with cooler weather, or iv) worsen with cooler weather. This class scheme assumed that temperature was the key driver of discussions; we found this to be representative of discussion on Twitter as well as the main

focus of academic literature on the weather's impact [4,5,7,8]. The inclusion, for instance, of both "improve with warmer weather" and "worsen with cooler weather" was to avoid any assumption of a linear effect of temperature given that non-linear effects have been documented [13]. From the set of 28,555 relevant tweets, a random sample of 2442 tweets (which included the pilot set of 200 tweets) was annotated per this scheme.

For qualitative analysis, the annotators recorded the mechanisms users reported for the weather's impact on coronavirus spread, such as sunlight destroying the virus. These mechanisms provided insight into the theories of the weather's impact being discussed and are reported in the Discussion.

3.1.2. Natural language processing and featurization

For our Effect Classifier, the same machine learning techniques were used from our Relevancy Classifier (as described above) with one modification: for the trinary classification, we optimized with respect to balanced accuracy, since AUC-PR and AUC-ROC do not extend to multiclass problems.

3.2. Analyzing tweets for topic via clustering

To extract unexpected topics of discussion, we performed unsupervised learning to cluster the tweets and determined topics through inspection of the clusters. After removing repeated tweets (not retweets) and attached article data, we used k -means clustering to group tweets into k clusters—other methods, specifically k -medoids and latent Dirichlet allocation [27] were also explored (see S7). Clustering was performed on the same TF-IDF vectors generated for effect analysis, and cluster sizes in $k = 10, 15, 20, 25,$ and 30 were tested. Each cluster was associated with an output of the top 20 keywords, based on highest TF-IDF scores. Outputs from each of the clustering configurations were inspected manually for the cohesiveness of topics.

4. Results

4.1. Data preparation and annotation

The data pipeline is displayed in Fig. 1, with inspiration taken from Ong et al. [28] As mentioned above, two rounds of annotation were performed. For relevancy classification, annotators labeled a random sample of 2786 tweets, and the Relevancy Classifier was trained on this. Then, for effect classification, the "effect" of a random sample of 2442 relevant tweets (out of 28,555) was annotated per the Effect Class and annotation scheme introduced earlier. Both sample sets were produced by uniformly sampling their respective parent set. That is, the relevancy sample set was selected from 84,201 tweets remaining after the rule-based sampling, and the effect sample set was selected from 28,555 tweets classified as relevant. By "uniformly" we mean that the number of tweets in a given month in the sample set is proportional to the number of tweets from that month in the parent set. The exact numbers were attained due to removing duplicate or unrelated tweets from each annotation set. Annotation results for the effect scheme are shown in Table 1; further, Table 2 describes the geographic breakdown by country of both the full set of collected tweets and the filtered set of relevant tweets.

Table 1
Manual Annotation Scheme for Effect and Class Proportions.

Class	Proportion (out of 2442)
Uncertain	40.4 % (987)
No Effect	33.5 % (817)
Effect	26.1 % (638)
Improve Warmer Weather	585
Worsen Warmer Weather	33
Improve Cooler Weather	4
Worsen Cooler Weather	16

Table 2
Number of Tweets from Countries Across Full Dataset and Relevant Tweets Set.

Full Dataset (166,005 tweets)		Relevant Tweets (28,555 tweets)	
United States	60,749	United States	9740
United Kingdom	13,579	United Kingdom	992
Canada	7159	India	912
India	4738	Canada	836
Australia	1739	Nigeria	446
Nigeria	1385	Pakistan	337
South Africa	1079	Australia	198
Ireland	1065	South Africa	142
Pakistan	866	Philippines	105
France	613	Kenya	100
Philippines	557	Germany	96
Germany	536	Spain	93
Kenya	529	Ireland	74
Other (<500 tweets)	10,143	Other (<73 tweets)	2068
No Data*	61,268	No Data*	12,416

* "No Data" represents tweets where neither the tweet nor tweet author had location data available.

4.2. Relevancy classification using machine learning

Our relevancy classifier identified tweets discussing the weather's impacts on COVID-19, with the volumes over time shown in Fig. 2. Four example peaks in activity are shown in the figure along with the most commonly shared headline in the dataset from that day (more details are available in S6). The best performing classifier for this phase of learning was Gradient Descent Support Vector Machine with TF-IDF featurization, with AUC-PR (95 % CI) = 0.862 (0.853, 0.871) and AUC-ROC (95 % CI) = 0.916 (0.907, 0.925).

4.3. Effect analysis

4.3.1. Manual annotation results

The 2442 annotated tweets were separated according to their effect label (effect, no effect, uncertain) and plotted in Fig. 3.

4.3.2. Machine learning classification results

Using the manual annotations for our Effect Classifier, we attempted to classify the perception of the effect that weather will have on COVID-19 transmission of a tweet according to the three effect classes. However, the multiclass scheme proved too difficult for machine to solve (see S5), but after collapsing our class scheme to a binary "effect" vs. "no effect/uncertain" (combining those two categories) the performance of the model improved (see Table 3 and S5). We still present these to show the machine did learn to identify effect to an extent, accomplishing our goal of identifying perception even after limiting our analysis to the coarser class scheme. The AUC-PR and AUC-ROC scores are reported in Table 3; for reference, a baseline classifier (one that randomly predicts the class) has an AUC-PR of 0.261—the proportion of the "effect" class in Table 1—and an AUC-ROC of 0.5. We note the presence of interesting dynamics in Fig. 3, showing how perception changed given that the sample was uniformly selected.

4.4. Clustering

The optimal configuration for k -means clustering was $k = 25$ to retrieve clear topics of discussion (see S7). After dropping 4803 repeated tweets, we clustered on 23,752 tweets. Twenty-four of the assigned clusters produced clearly delineated topics, while the remaining cluster was vague and contained general comments about weather and coronavirus. These clusters were not seeded with topics or themes, but rather are determined by the k -means algorithm.

Fig. 4 displays a heatmap tracking discussion frequency across ten selected topics over time. Boxes in the heatmap are shaded only for weeks where a topic exceeded its average level of discussion in the

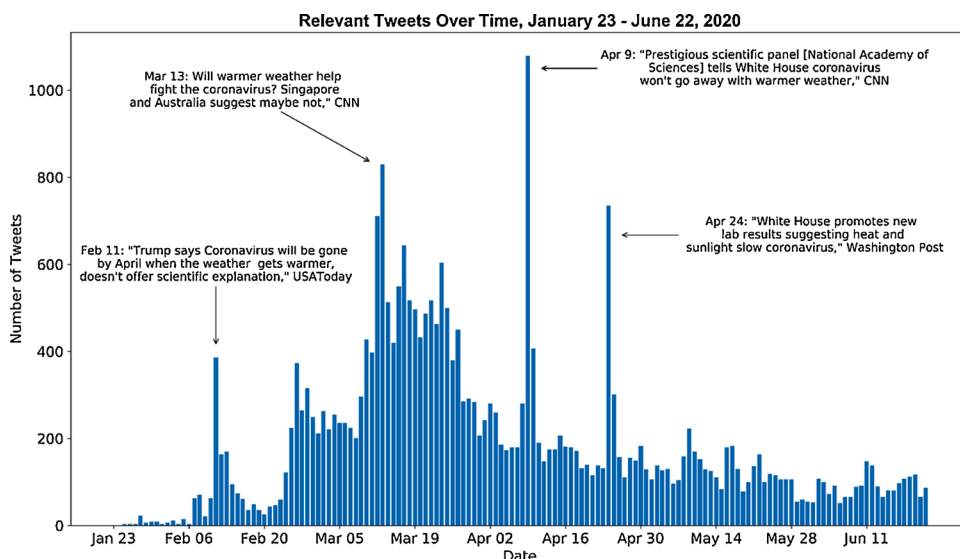


Fig. 2. Relevant original tweet volumes over time, with most frequent headlines and reporting organizations on four key peaks identified.

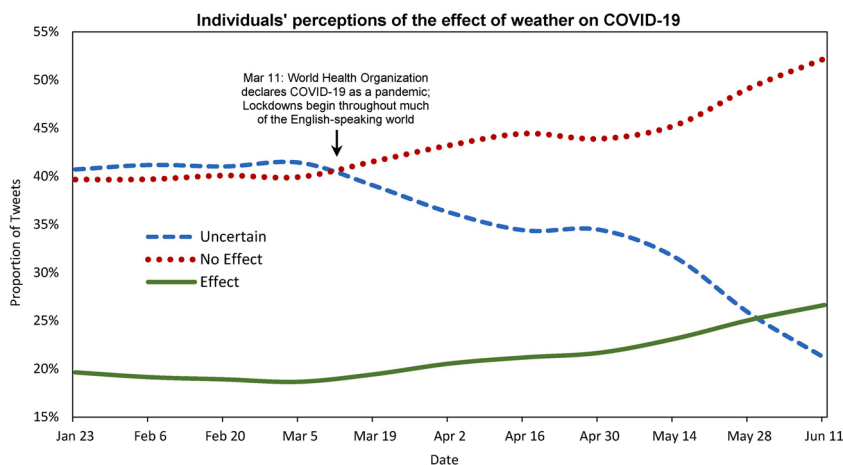


Fig. 3. Class proportion over time for annotated Tweets. Tweets are smoothed by 7 days, binned in 14-day windows, and weighted according to the individual tweet's number of retweets.

Table 3
Machine learning classification results.

Class	Proportion (out of 28,555)
No Effect/Uncertain	83.5 % (23,836)
Effect	16.5 % (4719)

Model: Gradient Descent Support Vector Machine, TD-IDF.

AUC-PR (95 % CI): 0.561 (0.542, 0.58).

AUC-ROC (95 % CI): 0.768 (0.749,0.787).

corpus, which allows for meaningful interpretation of when a topic is more active than usual.

The ten clusters plotted in Fig. 4 are particularly meaningful. Specifically, cluster 10 discussed the effect of cold weather on coronavirus spread; cluster 24 discussed the effect of hot weather on coronavirus spread; cluster 25 consisted of tweets proclaiming the relationship between different climates and general viral spread; cluster 11 discussed opinions propelled by scientific experts; cluster 4 focused on the ability of weather to 'kill' the coronavirus; clusters 5, 14, 18, and 21 referenced the Trump administration; cluster 6 included tweets comparing the coronavirus to influenza viruses; cluster 13 highlighted relationships between temperature and coronavirus spread; cluster 20 contained

tweets considering the ability of weather to 'slow spread' of the virus; and cluster 22 consisted of conversation revolving around social distancing. (See S7 for the top 20 keywords, sample tweets, and proportions for each cluster.) The appearance of several dedicated Trump clusters reflects the prominence of discussion surrounding Trump on the topic, in contrast to discussion surrounding other world leaders (see S7, Table S10 for comparison). Further, the geographic diversity of the dataset shown in Table 2 is also evident in the clustering algorithm output, as several specific geographic locations are discussed in the data in the context of warm or cold weather (see S7).

5. Discussion

Our analysis shows that Twitter user's perception of the weather's impact on the spread of COVID-19 varied greatly. Our results help quantify individuals' perceptions and reveal central topics of discussion surrounding weather and COVID-19 and have important implications for understanding where the public stands with respect to current public health knowledge on COVID-19.

From January through June 2020, the weather's impact on COVID-19 has been a present topic of discussion, where the volume of discussion ramped up between March 8 and April 1, coinciding with the

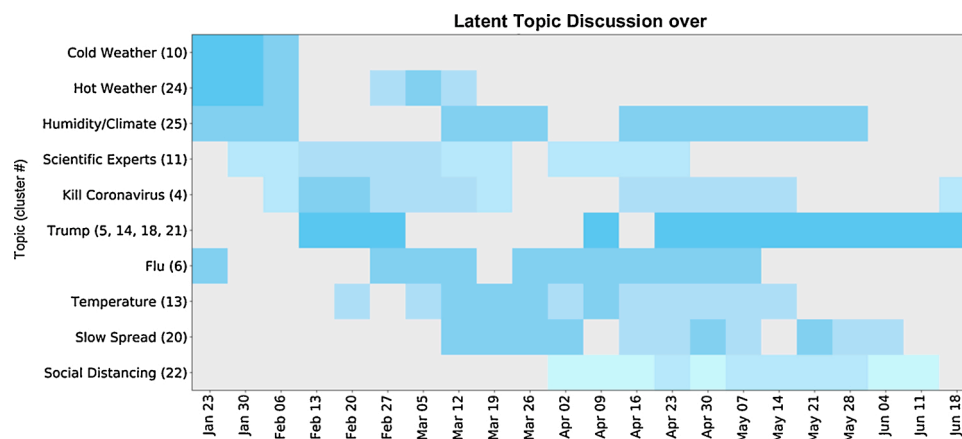


Fig. 4. Cluster Frequencies over Time by Week, color coding presents the frequency of discussion, where darker blue is the highest frequency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

beginning of stay-at-home orders throughout much of the world, including the United States and the United Kingdom. Furthermore, the spikes in the volume of discussion reflected significant events in the world. Fig. 2 documents four such events: Trump's comments in February claiming coronavirus would go away with the warm weather; [29] news coverage in March that Singapore and Australia suggest that the weather does not help fight the pandemic; [30] the National Academies of Science's response in early April to Trump's February claims; [31] and the White House's promotion in late April of lab results suggesting heat slows coronavirus [32]. This showed that Twitter conversation around the weather's impact on the spread of COVID-19 correlated with an increase in the spread of the virus and, inferably, impacted individuals' concerns [15].

Fig. 3 demonstrates a notable shift in opinion on the weather's impact through the progression of the pandemic, with significant movement beginning around March 11 (also coinciding the World Health Organization's declaration of COVID-19 as a pandemic, [33] as well as the beginning stay-at-home orders in the United States [34]). While there was a significant decrease in tweets displaying uncertain opinions, there was an increase in the proportion of tweets claiming no effect and in those claiming some effect of the weather on the spread of COVID-19. Similarly, the non-trivial proportion of tweets identified by the Effect Classifier claiming some effect is noteworthy given that the scientific community has not reached a clear consensus of the weather's impact on COVID-19 [35]. This claiming of an effect by users, regardless of whether it claims warming weather will improve or worsen the pandemic, shows that perception is formed as a result of broadcasted COVID-19 public health information and personal intuition on social media.

In Fig. 4, where cluster topic frequencies are plotted over time, trends are shown in discussions about the weather's impact on the spread of COVID-19. Furthermore, from January to February, there is a high frequency of discussion about cold weather and the flu, as these months exhibit both cold temperatures and the flu season, and the seasonality of COVID-19 was being discussed in reference to these topics. This was followed by an increase in discussion about reports made by scientific experts, from January 30 to March 19, about the weather's impact on the spread of COVID-19, as the virus was just beginning to spread globally, and its seasonal behavior was unknown. Simultaneously, there was an increase in discussion about Trump's comments from February 13–27, on April 9 and after April 23, following the same pattern seen in Fig. 2 where four of the illustrated peaks occurred. The high frequency of the Trump cluster shows the impact of President's statements and their constant relevance throughout the discussion of the weather's impact on the spread of COVID-19.

It is also interesting to note that the social distancing cluster did not

show up in Fig. 4 until April 2 and increased in frequency from May 7–28. This is likely because discussion about social distancing was not prevalent until after the lockdowns in much of the world in late March, and the discussion increased as the weather got warmer and people were more tempted to avoid social distancing guidelines. Similarly, discussion about social distancing peaked the same day discussion about Trump peaked on April 23, when in the US the White House promoted new evidence about heat possibly slowing the spread of COVID-19. This is curious, as many users claimed that heat will not slow the spread of COVID-19, only social distancing will.

Using clustering to reveal these topics helped understand which conversations were generating the greatest public response, allowing researchers to look into why these particular topics around the weather's impact on COVID-19 were standing out. The clustering analysis revealed a structure to the data beyond the effect class framework that we pursued for the supervised learning. For instance, comparisons of COVID-19 to the seasonality of influenza was a notable topic in size; sample tweets from that topic exhibited different claims of whether COVID-19 transmission will reduce in warmer weather like influenza (see S7 for sample tweets). It is important to take note of the possibility that some of the clusters may have existed outside our dataset of tweets—e.g., cluster topics such as the influenza virus or Trump could be discussed in connection to weather or COVID-19 in a realm outside of our study's purpose. Overall, our decision to include both our supervised and unsupervised analyses was verified by the different characteristics of the data revealed by each approach, which together enabled us to understand Twitter chatter.

During the manual annotation of tweets for effect, annotators recorded users' proposed mechanisms for the impact of weather, which are of interest as they exhibit potential misconceptions or unfounded theories. Some users who expected warm weather to decrease coronavirus spread discussed the following mechanisms: sunlight increasing Vitamin D levels and boosting immune response to the virus; hot weather destroying the viral capsid; and higher malaria resistance in populations with warmer climates correlating with resistance to COVID-19. Conversely, some users believed that warm weather could negatively impact the pandemic due to an increased temptation to avoid social distancing guidelines, increased transmission through air conditioning units or higher humidity, and decreased compliance to wear recommended personal protective equipment. These mechanisms demonstrate that in the absence of consensus among experts, speculative theories can take hold on social media. Understanding the drivers of this information can inform public health response to the pandemic. From an NLP perspective, automatically detecting causal mechanisms from text could be integrated into opinion mining to summarize perceptions more quickly [36]. Furthermore, given the variety of conflicting literature

published during the early phase of the pandemic, [4–12] work could be done to trace media coverage of these articles and understand how public opinion responded to press coverage of a given article.

This research is subject to limitations. As mentioned, the tri-class of “effect,” “no effect,” and “uncertain” problem proved too difficult for machine learning. Indeed, part of this arose from annotator difficulty in separating “no effect” and “uncertain” tweets. Several tweets were found to straddle the border of these two categories, partially due to the similarity of words across the “no effect” and “uncertain” tweets. This partly explains why collapsing these two categories into one improved our analysis performance enough to present results, and our adjusted Effect Classifier was able to successfully recognize users who claimed an effect.

An additional limitation in the effect annotation scheme was that we did not label for the magnitude of the effect. With this, we lose the nuance of whether tweets are claiming a strong, impactful or weak, inconsequential effect of the weather. One solution to this is to annotate for ‘weak’ or ‘strong’ effect or assign a numerical score for the strength of effect; with more ample training data it is plausible a model may successfully learn which tweets claim a strong effect or otherwise.

One significant language pattern that helped train our NLP analysis was the use of certain geographical locations to support a claim. For example, annotators noticed that warm locations, such as Florida and Singapore, were typically mentioned amongst users as a counterexample to undermine the possibility that warm weather will reduce the spread of COVID-19, and the names of these locations became a negative predictor for the “effect” class. Of course, not all mentions of warm locations in the data were as part of a counterexample, which exhibits one limitation of our model. Additionally, the Effect Classifier found the mention of “Trump” to be an accurate predictor for the “no effect/uncertain” class; this was largely due to sarcastic responses to Trump’s February predictions of the weather’s impact. Future directions include improving the performance of the Effect Classifier to detect more nuances of language, such as sarcasm and tone, which confused our models in some instances and are well-documented as difficult for machine learning models [37].

6. Conclusion

Our analyses revealed a surprising variety in conversations discussing potential seasonal impacts on COVID-19. The discussion went beyond the effect framework we chose that was centered around temperature and revealed various indirect impacts of weather as well, such as warming weather tempting the public to violate social distancing guidelines and hence accelerating the spread of COVID-19. Similarly, the presence of unsupported theories such as increased air-conditioning use during warmer months worsening spread or increased transmission through mosquitos raises questions of how many subscribe to them. With these results in mind, social media can be used to crowdsource such mechanisms and provide topics for study in order to address public misconceptions. Especially during a pandemic, when everything is novel and unsettling for most, the understanding of public opinion is crucial for public health. In the future, computational methods could be used to detect public’s opinion in real-time from social media to prepare for pandemic responses. Additionally, there is room to implement methods to measure and understand opinion in public health contexts, as well as understand how media coverage of studies percolate through the public. This study demonstrated a means to identify misconceptions in the general public, and the specific misconceptions encountered show that more work needs to be done to educate the public and correct these misunderstandings.

Funding

No funding was used to conduct this study.

Author contributions

MG and BJ conducted pilot testing for data collection, and MG and AB worked jointly on final data collection, preparation, and machine learning classification analyses. JR, BJ, and MG annotated training data and contributed to qualitative analyses of data. MG designed the topic analysis, for which BJ, AO, AB, and MG wrote code and BJ executed. AO conducted validation testing reported in the supplementary materials. JR led the drafting of the manuscript with assistance from BJ, MG, AB, and MSJ. MSJ conceived the study, supervised the project, and revised the manuscript for important intellectual content.

Summary Points

- o The weather’s potential ability to curb coronavirus spread has been a key topic since the start of the pandemic.
- o Twitter data have been studied during past public health events to understand public perception and inform response.
- o We employed text mining approaches to examine how individuals perceive the impact of weather on COVID-19 and how their understanding evolves throughout the pandemic.
- o Suggested effects of weather exhibit a disconnect between scientific knowledge and public perception.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

We thank Yicheng Wang, Elizabeth Mason, and Heresh Amini who provided feedback and suggestions. We also thank Catherine DiGennaro for her contributions in initiating the research.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijmedinf.2020.104340>.

References

- [1] J. Shaman, E. Goldstein, M. Lipsitch, Absolute humidity and pandemic versus epidemic influenza, *Am. J. Epidemiol.* 173 (2) (2010) 127–135.
- [2] A. Altamimi, A. Ahmed, Climate factors and incidence of Middle East respiratory syndrome coronavirus, *J. Infect. Public Health* (2019). In press.
- [3] J. Yuan, et al., A climatologic investigation of the SARS-CoV outbreak in Beijing, China, *Am. J. Infect. Control* 34 (4) (2006) 234–236.
- [4] A. Notari, Temperature dependence of COVID-19 transmission, *medRxiv* (2020) 2020, 03.26.20044529.
- [5] G.F. Ficitola, D. Rubolini, Climate affects global patterns of COVID-19 early outbreak dynamics, *medRxiv* (2020) 2020, 03.23.20040501.
- [6] J. Bu, et al., Analysis of meteorological conditions and prediction of epidemic trend of 2019-nCoV infection in 2020, *medRxiv* (2020) 2020, 02.13.20022715.
- [7] Q. Li, et al., Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, *N. Engl. J. Med.* 382 (13) (2020) 1199–1207.
- [8] C. Merow, M.C. Urban, Seasonality and uncertainty in COVID-19 growth rates, *medRxiv* (2020) 2020, 04.19.20071951.
- [9] W. Luo, et al., The role of absolute humidity on transmission rates of the COVID-19 outbreak, *medRxiv* (2020).
- [10] N. Islam, S. Shabnam, A.M. Erzurumluoglu, Temperature, humidity, and wind speed are associated with lower Covid-19 incidence, *medRxiv* (2020) 2020, 03.27.20045658.
- [11] B. Oliveiros, et al., Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases, *medRxiv* (2020) 2020, 03.05.20031872.
- [12] M.M. Sajadi, et al., Temperature, Humidity and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19, Preprint, 2020.
- [13] R. Xu, et al., The modest impact of weather and air pollution on COVID-19 transmission, *medRxiv* (2020) 2020, 05.05.20092627.
- [14] L. Lin, et al., Media use and communication inequalities in a public health emergency: a case study of 2009–2010 pandemic influenza A virus subtype H1N1, *Public Health Rep.* 129 (6 suppl4) (2014) 49–60.
- [15] L. Singh, et al., A First Look at COVID-19 Information and Misinformation Sharing on Twitter, 2020.

- [16] M. Le Page, Will heat kill the coronavirus? *New Sci.* 245 (3270) (2020) 6–7.
- [17] Q.B.Y. Jameel, Will Coronavirus Pandemic Diminish by Summer? Elsevier BV, 2020, p. 15.
- [18] A. Culotta, Towards detecting influenza epidemics by analyzing twitter messages, *Proceedings of the First Workshop on Social media Analytics* (2010).
- [19] L. Hong, B.D. Davison, Empirical study of topic modeling in twitter, *Proceedings of the First Workshop on Social media Analytics* (2010).
- [20] A. Sarker, A. DeRoos, J. Perrone, Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework, *J. Am. Med. Inform. Assoc.* 27 (2) (2019) 315–329.
- [21] M.E. Peters, et al., Deep contextualized word representations, arXiv preprint (2018) arXiv:1802.05365.
- [22] P.D. Turney, P. Pantel, From frequency to meaning: vector space models of semantics, *J. Artif. Int. Res.* 37 (1) (2010) 141–188.
- [23] F. Pedregosa, et al., Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [24] J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves, in: *Proceedings of the 23rd International Conference on Machine Learning, Association for Computing Machinery: Pittsburgh, Pennsylvania, USA, 2006*, pp. 233–240.
- [25] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, *PLoS One* 10 (3) (2015) e0118432.
- [26] L.A. Jeni, J.F. Cohn, F.D.L. Torre, Facing imbalanced data—recommendations for the use of performance metrics, 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (2013).
- [27] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, *J. Mach. Learn. Res.* 3 (2003) 993–1022. Jan.
- [28] C.J. Ong, et al., Machine learning and natural language processing methods to identify ischemic stroke, acuity and location from radiology reports, *PLoS One* 15 (6) (2020) e0234908.
- [29] C. Subramanian, S. Behrmann, D. Jackson, Trump Says Coronavirus Will Be Gone by April When the Weather Gets Warmer, Doesn't Offer Scientific Explanation, in *USA TODAY*, 2020.
- [30] Griffiths, J. Will warmer weather help fight the coronavirus? Singapore and Australia suggest maybe not. *CNN* 2020; Available from: <https://edition.cnn.com/2020/03/12/asia/coronavirus-flu-weather-temperature-intl-hnk/index.html>.
- [31] National Academies of Sciences Engineering and Medicine, Rapid Expert Consultation on SARS-CoV-2 Laboratory Testing for the COVID-19 Pandemic (April 8, 2020), in *Rapid Expert Consultations on the COVID-19 Pandemic: March 14, 2020–April 8, 2020*, National Academies Press, US, 2020.
- [32] A. Freedman, J. Samenow, White House Promotes New Lab Results Suggesting Heat and Sunlight Slow Coronavirus, *The Washington Post*, 2020.
- [33] World Health Organization, WHO Director-General's Opening Remarks at the Media Briefing on COVID-19-11 March 2020, 2020.
- [34] White House, Proclamation on Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID-19) Outbreak, White House, 2020.
- [35] E. Cohen, Prestigious Scientific Panel Tells White House Coronavirus Won't Go Away With Warmer Weather, *CNN*, 2020.
- [36] E. Nazaruka, An overview of ways of discovering cause-effect relations in text by using natural language processing, in: *International Conference on Evaluation of Novel Approaches to Software Engineering*, Springer, 2019.
- [37] B. Pang, L. Lee, Opinion mining and sentiment analysis, *Found. Trends® Inf. Retrieval* 2 (1–2) (2008) 1–135.