Learning from Tweets: Opportunities and Challenges to Inform Policy Making During Dengue Epidemic

FARHANA SHAHID, Cornell University
SHAHINUL HOQUE ONY, Brac University
TAKRIM RAHMAN ALBI, Brac University
SRIRAM CHELLAPPAN, University of South Florida
ADITYA VASHISTHA, Cornell University
A. B. M. ALIM AL ISLAM, Bangladesh University of Engineering and Technology

Social media platforms are widely used by people to report, access, and share information during outbreaks and epidemics. Although government agencies and healthcare institutions in developed regions are increasingly relying on social media to develop epidemic forecasts and outbreak response, there is a limited understanding of how people in developing regions interact on social media during outbreaks and what useful insights this dataset could offer during public health crises. In this work, we examined 28,688 tweets to identify public health issues during dengue epidemic in Bangladesh and found several insights, such as irregularities in dengue diagnosis and treatment, shortage of blood supply for Rh negative blood groups, and high local transmission of dengue during Eid-ul-Adha, that impact disease preparedness and outbreak response. We discuss the opportunities and challenges in analyzing tweets and outline how government agencies and healthcare institutions can use social media health data to inform policy making during public health crises.

CCS Concepts: • **Human-centered computing** → **Social networking sites**.

Additional Key Words and Phrases: public health crisis; dengue; epidemic; outbreak; Twitter; blood donation; health policy; HCI4D; Bangladesh

ACM Reference Format:
Farhana Shahid, Shahinul Hoque Ony, Takrim Rahman Albi, Sriram Chellappan, Aditya Vashistha, and A. B. M. Alim Al Islam. 2020. Learning from Tweets: Opportunities and Challenges to Inform Policy Making During Dengue Epidemic . In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 4, CSCW1, Article 65 (May 2020), 27 pages. https://doi.org/10.1145/3392875

1 INTRODUCTION

Governments and healthcare organizations worldwide are increasingly using social media to gain insights about various infectious diseases, such as influenza [8, 49, 59], Zika virus disease [71, 97], dengue [61], and Ebola [38, 41]. Most previous and current efforts have utilized social media data for early prediction of disease outbreaks [7, 14, 17]. However, the early epidemic forecast models are rarely evaluated during or after the event due to lack of proper guidance, communication, and consultation among different stakeholders involved in outbreak management [30, 39].

Authors' addresses: Farhana Shahid, Cornell University, fs468@cornell.edu; Shahinul Hoque Ony, Brac University, shahinul.hoque.ony@g.bracu.ac.bd; Takrim Rahman Albi, Brac University, takrim.rahman.albi@g.bracu.ac.bd; Sriram Chellappan, University of South Florida, sriramc@usf.edu; Aditya Vashistha, Cornell University, adityav@cornell.edu; A. B. M. Alim Al Islam, Bangladesh University of Engineering and Technology, alim_razi@cse.buet.ac.bd.

https://doi.org/10.1145/3392875

Instead of developing epidemic forecast models, we aim to explore how social media can provide useful insights for disease surveillance, outbreak response management, and public health policy as outbreaks unfold. Although HCI community has well-studied the use of social media during humanitarian crises [64, 117], natural calamities [76, 100, 101], and man-made disasters [21, 48], there is a dearth of research on how people interact on social media during the critical moments of disease outbreaks and epidemics [44]. This shortage is more acute in low- and middle-income countries with emergent social media practices and crippled public health systems.

To address the existing gaps in literature and examine how social media can inform policy making during public health crises, we take the case of recurrent dengue outbreaks in Bangladesh. In particular, we focus our attention on the 2019 dengue epidemic in Bangladesh which resulted in over 101,000 infections and 179 deaths, making it the worst dengue epidemic in the last 20 years [115]. The small-scale government surveys failed to capture the urgency of the situation as they reported only sporadic cases of dengue within the capital [84, 98]. Moreover, poor quality of government health data and lack of data accessibility made it difficult to elicit meaningful and actionable insights to control transmission and manage the epidemic [30, 114]. This brings us to question how government agencies and healthcare institutions can use social media platforms, such as Twitter, to discover insights that inform policy making during dengue outbreaks? Specifically, we sought to answer the following research questions:

RQ1: What can dengue related tweets *reliably* tell us about the public health issues in Bangladesh?

RQ2: How Twitter data can *guide* public health policy for dengue epidemic in Bangladesh?

To answer our research questions, we extracted 28, 688 dengue related tweets in Bangladesh and developed biterm topic models (BTM) to discover the topics that captured public attention. We also compared the insights obtained from tweets with the data present in government reports and examined how these two datasets complement each other. In response to **RQ1**, we found several interesting insights about public health issues in Bangladesh. For example, many people shared mismanagement in dengue outbreak response, such as lack of testing facilities and high costs of treatment for dengue, on Twitter. The insights obtained by analyzing tweets indicated shortcomings in government-reported dengue statistics. We also discovered interesting patterns, for example, more deaths of women, new outbreaks around Eid-ul-Adha, and higher demand for Rh negative blood groups for platelet transfusion of dengue patients, during the epidemic. In response to **RQ2**, we provide a set of recommendations on how government agencies and healthcare institutions can leverage social media health data to improve disease surveillance, monitor policy violations, capture early-stage infections, take proactive measures for expected outbreaks, and streamline outbreak response. In this work, we make three contributions. First, we demonstrate the usefulness of Twitter in discovering public health issues and uncovering useful insights during an epidemic. Second, we show how various social, cultural, geographical, and epidemiological factors impact response to dengue outbreak in Bangladesh. Third, we provide actionable recommendations on how government agencies and healthcare institutions can use Twitter to inform health policy during an epidemic.

In the paper that follows, we first discuss related work at the intersection of social media, disease surveillance, and public policy. We then describe our datasets, quantitative methods, and findings. Our discussion offers pragmatic challenges in analyzing tweets and recommends how government agencies and healthcare institutions can tap on Twitter to improve outbreak detection and mitigation strategies.

2 RELATED WORK

We now situate our research in a body of related work examining how social media platforms and human-centered approaches are used for improving disease surveillance and informing public policy during a public health crisis.

2.1 Social Media based Disease Surveillance

Early prediction of outbreaks, based on web and social media data, is well-studied for various diseases [8, 25, 72, 106]. For example, Chunara et al. [27] used data from news media to study the epidemiological pattern of the 2010 Haitian Cholera outbreak. Yuan et al. [123] developed an Influenza epidemic monitoring system based on the search queries from Baidu. Even in case of dengue, several models have been developed to monitor and forecast the epidemic based on Google search queries [22] as well as Wikipedia access logs and category links [90].

Apart from web, many researchers have utilized social media data as *sensors* to provide early warnings of disease outbreaks [7, 58]. For example, researchers have used geo-tagged tweets to model the spread of infectious diseases [56, 60, 80, 92, 102]. However, the acute shortage of geo-tagged tweets [56, 107] as well as associated security and privacy risks make this approach less appealing [95]. In lieu of geo-tagged tweets, researchers have also relied on user-provided locations during registration as a proxy for geo-tags [67]. However, user-provided locations often have high variance due to human movements and are too high-level to be of use [99]. In addition to challenges in building these models, researchers have also struggled in evaluating these models due to a lack of proper guidelines and collaboration among different stakeholders [30, 39]. While more research is required to build better epidemic forecast models and evaluation techniques, in this work, our goal is to elicit new insights from dengue related public tweets posted during an epidemic.

For example, prior work have examined how human movements due to socio-cultural norms amplify transmission of infectious diseases [18, 40, 66, 104, 118], often by relying on cell phone records [20, 65], geo-tagged tweets [60], or search queries [122]. Our work contributes to this line of research by conducting spatio-temporal analysis of dengue cases reported in tweets to examine how large-scale human movements during Eid impact dengue outbreak.

2.2 Human-Centered Approaches to Examine Public Health Crisis

Epidemiological methods for monitoring disease outbreaks often fail to capture social nuances, affective relationships, and other human-centered factors and concerns [11]. Recently, several researchers have used a human-centered lens to design interventions for predicting and analyzing disease transmission. For example, Ghosh et al. [41] developed an automated tweet categorization system to fulfill the information needs of different stakeholders during an epidemic.

Many researchers have also explored how social media platforms are used by people during public health crises. For example, Kannan et al. examined how Malaysians use social media to obtain dengue related information [54]. Gui et al. investigated how individuals turned to social media during Zika outbreak for information gathering, social learning, and travel-related decision making [44]. Some researchers have also designed new social media applications to enable people to access, report, and share information. For example, Lwin et al. designed “Mo-buzz”, a social media application which predicts dengue outbreaks and allow users to post pictures of breeding sites of mosquitoes [62]. In addition, some researchers have used Twitter to analyze public sentiments, frustrations, and opinions related to disease treatment and outbreak mitigation plans [43, 74, 96]. Our work extends this body of research by examining how people in Bangladesh use Twitter to share information and highlight shortcomings in dengue outbreak response.

Prior work have also examined how people use social media platforms to organize resources and cope with public health crisis. For example, social media users are increasingly using these platforms to post and share messages requesting blood donations [1, 6]. In response to this emergent behavior, Mathur et al. developed a model to classify the tweets requesting urgent blood donation to improve discoverability and reach of these time-sensitive requests [69]. Since the demand for platelet and fresh frozen plasma rises exceptionally high during dengue outbreaks [108], we explore how demand varies across different blood groups by analyzing personal blood donation requests on Twitter during dengue epidemic in Bangladesh.

Recently, several researchers have studied the propagation of rumors and misinformation on social media platforms during outbreaks of infectious diseases, such as Ebola virus disease [38, 79, 86], Zika virus disease [42, 97], and H1N1 influenza [26]. For example, Luque and Bau studied how anti-vaccine activists manipulated information on social media questioning the efficacy and safety of a vaccine [35]. Such rumors can influence public confidence in vaccines [55] and severely undermine public health policies [35]. Because of concerns around propagation of misinformation and rumors during outbreaks, public health specialists have mixed feelings about the potential of social media platforms in informing public health policy [52]. Although we do not directly contribute to the discourse on social media misinformation during disease outbreaks, our work reports the reluctance of government agencies and healthcare institutions to engage with social media.

2.3 Social Media and Health Policy

Despite the presence of rumors and misinformation, many studies have highlighted the potential of social media in supporting and improving public health infrastructure [25, 93, 109]. For example, Yeung argued that social media find diverse pathways to understand public health attitudes and identify key cultural components to promote health and well-being [121]. Charalambous discussed how social media can influence cancer policy by disseminating information, shaping public opinion, encouraging participation, and establishing connection among different stakeholders [24]. Broniatowski et al. studied how Twitter-based influenza surveillance can provide actionable information to policymakers working at the municipal, national, and health agencies [19]. Yaya et al. utilized social media to empower women with health related knowledge to prove the utility of these platforms in making health policy related to malaria elimination program [120]. Our work extends this literature by showing how policy makers can use tweets to improve disease surveillance, monitor policy violations, discover new outbreaks, and streamline outbreak response.

Prior studies have underscored the importance of understanding how different factors and their relationships impact sustainability and efficacy of policies to manage dengue outbreaks. For example, Fonseca and Zicker conducted a longitudinal evaluation of dengue research networks and found structured and organized collaborative efforts between central institutions to be prominent in the success of dengue response efforts [29]. Spiegel et al. advocated integrating ecological and societal complexities as well as perspectives of various stakeholders to improve the prevention and control efforts [103]. Our work contributes to this literature by examining social, cultural, geographical, and epidemiological dimensions of dengue epidemic in Bangladesh and their effect on outbreak response efforts. Our work provides actionable recommendations on how government agencies and healthcare institutions can use health information and reports on social media to strengthen health infrastructure in Bangladesh.

3 METHODOLOGY

We now present our methods to analyze dengue outbreaks in Bangladesh. We first curated multiple datasets to analyze public responses on Twitter and reports published by the government. We used

Table 1. Different keywords and locations used to retrieve dengue related tweets in the context of Bangladesh.

Keywords		Location
dengue	#fever	Dhaka
aedes	#blood	Chittagong
aegypti	#hospitalized	Khulna
#bangladesh	#dead	Barisal
#denguefever	#infected	Sylhet
#mosquito	#prevention	Rajshahi
#epidemic	#infection	Rangpur
#outbreak	#treatment	Mymensingh
#dengueoutbreak	#hospital	. . .

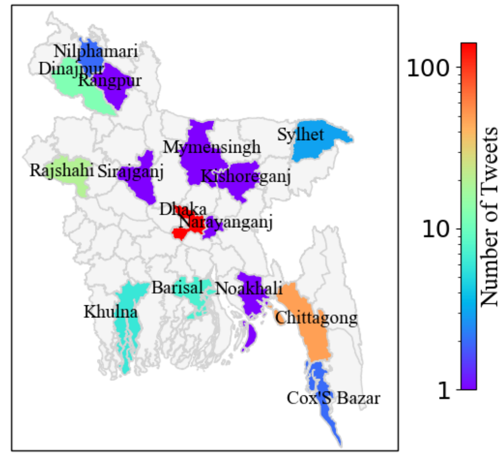


Fig. 1. Local distribution of 268 geo-tagged tweets related to dengue in Bangladesh.

topic modeling to identify the breadth of topics present in the dengue related public tweets. We then conducted different statistical analyses on the data available from tweets and government reports to examine how much they complement each other.

3.1 Datasets: Data Collection and Pre-processing

3.1.1 Twitter. We used TweetScraper [91] to curate dengue related public tweets published from July 2010 to September 2019. We chose three initial keywords (dengue, aedes, and aegypti) as used in the previous work [7, 43, 67, 80] and performed one-step snowball sampling following the method described in [73, 78] to find more hashtags. We included all the hashtags that were relevant to dengue (as rated by three medical students) and had frequency > average frequency of all the hashtags. We then paired each keyword with *dengue* (e.g., ⟨dengue, #outbreak⟩, . . .) to retrieve dengue related tweets. We also used the names of 8 divisions and 64 districts with these keywords to filter the tweets relevant to the incidences of dengue in Bangladesh. Table 1 shows the list of keywords and locations we used in our search queries.

We collected 28,688 tweets (excluding retweets) related to dengue outbreak in Bangladesh. Nearly 80% of these tweets were generated in 2019, the year Bangladesh faced its worst and most dangerous dengue epidemic [36]. Our dataset is much smaller than similar datasets curated in Brazil [43, 67, 73], Philippines [28], and Indonesia [92]. This is because Bangladesh has much lower Internet penetration and is a lower tweet-density area [53, 107].

We found 354 tweets (1.23% of tweets in our sample) with geo-tags, comparable to the proportion found in other studies [56, 107]. Nearly 75% of these tweets were generated in Bangladesh and 53% were generated in the capital city Dhaka. Figure 1 plots geographic distribution of these tweets, indicating availability of tweets from all eight divisions of Bangladesh. Due to low availability of geo-tagged tweets, we used LNEx [3, 4] to extract location names from the text of our tweets. We found 6, 204 (21.6%) tweets that mentioned a location in Bangladesh. Using the list of hospitals in Bangladesh [32, 116], we also extracted tweets that mentioned healthcare institutions. We found 31 hospitals (25 public, 6 private) listed in our tweets, which we used for analysis in Section 4.2.

Next, we used R toolkit UDPipe to identify the reported cases of dengue from the tweets. We found 1, 778 (6.2%) tweets that report the number of people infected, hospitalized, or dead due to dengue. We observed that multiple tweets mentioned the same cases of dengue. For example, consider the following two tweets with different sentence structures: (1) *"Bangladesh #dengue cases reach nearly* 25, 000 *as nationwide outbreak continues unabated"* and (2) *"Dengue outbreak in Bangladesh continues unabated. Number of people infected with dengue almost* 25000." Both tweets were posted on the same day by different users. To prevent double counting, we measured similarity among all pairs of tweets by using Jaccard similarity [45], and counted only once that reported the same incidence. We considered two tweets to report the same incidence if their Jaccard similarity is greater than 0.5, they report same number of dengue cases, and they are posted within a 24-hour interval. Finally, we obtained 1,440 tweets that were used with the location names extracted by LNEx for spatio-temporal analysis presented in Section 4.2 and 4.4. Moreover, we used Twitter API to extract profile descriptions of 7, 480 unique Twitter users in our dataset to assess their roles during dengue epidemic as reported in Section 4.5.

Since many dengue patients undergo blood transfusion, we used regular expressions [1] to find tweets that mentioned blood groups and the amount of blood required for dengue patients. We found 1, 105 (3.9%) tweets that requested for voluntary blood donation for dengue patients in Bangladesh. We also examined which blood groups were in high demand during dengue epidemic compared to the non-epidemic period. Using TweetScraper and blood groups as keywords[1], we found 95, 170 public tweets looking for volunteer blood donors during non-epidemic period (from 2008–2019). Of these, 642 (0.6%) tweets originated from Bangladesh or listed hospitals in Bangladesh. We used these two sets of tweets (i.e., 1, 105 tweets during dengue epidemic and 642 tweets during non-epidemic period) for our analysis on blood donation requests presented in Section 4.6.

There is an active debate on ethics of analyzing publicly available Twitter data [111]. For example, Zimmerman argued that publicly posted tweets may not imply consent to analyze these tweets [124]. Maddock et al. outlined ethical challenges in analyzing deleted tweets, since deletion indicates participant's desire for those tweets to be forgotten [63]. Other researchers have suggested best practices when working with such publicly available data. For example, Bruckman advocated making slight modifications in verbatim content to make it difficult to tie it back to content author [51]. In our IRB-approved study, we followed such best practices to protect the identity of Twitter users in our sample. For example, we made slight modifications to example tweets presented in our paper. We also replaced all personal identifiable information, such as phone numbers, addresses, and names, with pseudonyms before analyzing the datasets. Although we wanted to discard deleted tweets in our dataset, we struggled to identify such tweets because of technical complexity and lack of required resources [63].

3.1.2 Government Reported Dengue Statistics. We used government reported dengue statistics to compare the insights obtained from analyzing Twitter data. While Directorate General of Health Services (DGHS) in Bangladesh publishes daily dengue statistics [31], we could collect reports only from August 2019–September 2019 . This is because DGHS website often discards old reports while publishing the latest ones. From these reports, we collected the number of people hospitalized and dead due to dengue both at district and hospital levels. These health bulletins report dengue cases for 8 public and 18 private hospitals in Bangladesh.

We also collected monthly dengue status reports from Institute of Epidemiology, Disease Control and Research (IEDCR) in Bangladesh that report total monthly cases of dengue for the whole country from 2008–2019 [83]. However, for our research, we considered the most recent time frame from 2015–2019, as this period has been least studied in contemporary research work [33, 50, 75, 77]. Besides, 98% of our dengue related tweets are from 2015–2019.

[1]ABO blood group: A, B, AB, O. Rh antigen: +, -, +v, -v, +ve, -ve, pos, neg, positive, negative (each antigen keyword was used as-is, within () and [], e.g., +, (+), [+].)

Table 2. Overview of our collected datasets.

Data	n (%)
Dengue related tweets in Bangladesh (28,688) 2010–2019	Geo-tagged tweets: 354 (1.23%) Location names: 6,204 (21.6%) Dengue cases: 1,440 (5.02%) Blood donation: 1,105 (3.9%)
Blood donation tweets from non-epidemic period (95,170) 2008–2019	Blood donation requests in Bangladesh: 642 (0.6%)
Daily dengue statistics from DGHS, Bangladesh Aug–Sep, 2019	Dengue cases at district and hospital levels
Monthly dengue reports from IEDCR, Bangladesh 2008–2019	Monthly total dengue cases for the whole country

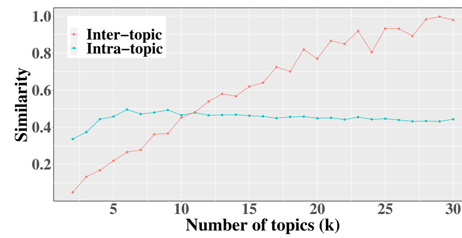


Fig. 2. Intra-topic and inter-topic similarity for different number of topics obtained using BTM.

We also collected data from two DGHS surveys conducted to measure the density of Aedes vector population in different areas of Dhaka city [81, 82]. In the first phase, DGHS inspected 100 sites of 98 wards in 2 city corporations of the capital [81]. In the second phase, they surveyed 14 different public places in Dhaka from July 31–August 4, 2019 [82]. We combined data from these two surveys and compared it with our Twitter dataset for the analysis presented in Section 4.4. Table 2 summarizes details of the datasets used in our analyses.

3.2 Analyses

We first examined which aspects of the dengue epidemic and its consequences were at the forefront of public attention. We then conducted additional analyses on these emergent themes to identify how dengue related tweets provide additional insights to inform public health policy.

3.2.1 Topic Modeling. To discover dominant themes from dengue related tweets, we used biterm topic model (BTM) [119]. We chose BTM instead of conventional topic models (e.g., LDA [16] and PLSA [47]) since it works well on the sparse short texts of tweets [119]. We did not use Ailment Topic Aspect Model as it is more suitable to uncover aspects across multiple diseases [87, 89].

We first tokenized the tweets and annotated each token with its corresponding universal parts of speech using UDPipe. After removing all the stop words, punctuation, and other symbols, we built BTM to extract different topics from dengue related tweets. We developed multiple biterm topic models for different values of the number of topics (k). For each model, we set symmetric dirichlet prior probability of a topic, $\alpha = 50/k$ and symmetric dirichlet prior probability of a word given the topic, $\beta = 0.01$. We chose these values of α and β as they have been used previously to build aggregated topic models to increase social media topic coherence [15] and to mine health topics from Twitter [88]. We performed 4,000 iterations of Gibbs sampling to build each model.

We chose the number of topics (k) for our model based on intra-topic and inter-topic similarity, since these measures have been used previously to assess topic quality in tweets while developing dengue outbreak detection models [73]. For each model, high intra-topic similarity indicates good topic coherence and greater inter-topic similarity indicates poor separation. Therefore, for optimal number of topics we need both high intra-topic similarity (good coherence) and low inter-topic similarity (good separation). Figure 2 shows that intra-topic similarity is highest (0.336) and inter-topic similarity is low (0.26) at $k = 6$. Therefore, we used six topics to explain emergent themes on dengue related tweets in Bangladesh.

3.2.2 Statistical Analyses. To answer RQ1, we performed different statistical analyses on the emergent themes and examined how insights from tweets complement the data obtained from government reports. We used parametric tests for data that follow normal distribution and performed non-parametric tests otherwise. To control false discovery rate for multiple hypothesis testing, we applied Benjamini-Hochberg error correction [70] on all the results. Finally, to answer RQ2, we examined how these insights can inform policy making approaches for dengue epidemic in Bangladesh by efficient real-time reporting of dengue prevention and mitigation efforts.

4 FINDINGS

Our analysis of dengue related public tweets in Bangladesh identified six major topics. Table 3 shows these topics, top 20 keywords, example tweets, and frequency of tweets measured using the method described in [51]. We now examine each topic closely and analyze social, cultural, geographical, and epidemiological aspects of dengue epidemic in Bangladesh to answer our RQ1.

4.1 Incidences of Dengue

Nearly 16% tweets in our sample discussed incidences of dengue. People tweeted about being infected with dengue and posted the sufferings of family and friends due to dengue fever. We also found many educational tweets where people posted infographics about causes, symptoms, and treatment of dengue. People also shared resources, for example, android applications for dengue information, with each other. Many tweets contained memes or mnemonic devices to make it easier for others to recognize symptoms. These findings indicate that Twitter quickly emerged as a platform for peer-training and peer-sharing of health information.

> *"#Dengue symptoms have changed. Doctors say the virus of this fever is stronger now. There are 4 types of dengue virus. To date, people used to be infected with virus one but nowadays they are being infected with virus two and three . . . "*

> *"Symptoms, prevention, and treatment of dengue fever: apk download for Android - http://***"*

We found a surprising trend related to deaths reported on Twitter. Of the 777 tweets reporting dengue related deaths, 109 (14%) mentioned deaths of women, 20 (2.6%) of men, 26 (3.3%) of children, 15 (1.9%) of girls, 1 (0.1%) of boys, and 35 (4.5%) of doctors. This shows six times more deaths of females than males. While our dataset did not reveal reasons for this unexpected trend, we corroborated this finding with media reports that confirmed more deaths of women during 2019's dengue epidemic [68].

Next, we examined how the deaths reported by the government agencies compare with the deaths reported on Twitter. For this analysis, we used the DGHS dataset and compared the number of deaths from January 2019 to September 2019. DGHS reported 100 deaths in 21 districts, whereas tweets indicated 67 deaths in 23 districts. We found a medium size correlation between deaths reported by the government and by Twitter users ($r_\tau = 0.40$, $p < .01$). Figure 3a and 3b plot geographic distribution of reported deaths in different districts in Bangladesh. Compared to the government data, Twitter data reported more deaths in five districts and non-zero deaths in eight new districts, indicating its potential as a tool to reflect ground realities. One tweet mentioned:

> *"Red tape hides actual #dengue death figures in #Bangladesh http://www.newagebd.net/article/82679/red-tape-hides-actual-dengue-death-figures"*

We also found some obvious insights, for example, high number of deaths in densely populated cities like Dhaka, suggesting that the government should strengthen its prevention and mitigation efforts in such regions. Some tweets also discussed how external factors, such as floods and poor infrastructure, deteriorated government's response to dengue outbreak. Twitter users expressed

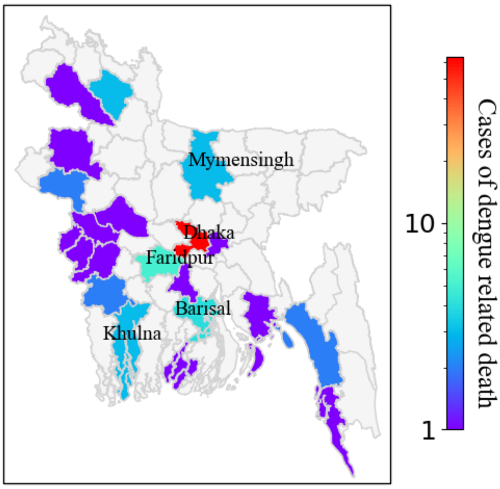
Table 3. Six topics from biterm topic model (BTM) of dengue related tweets along with relative topic frequency, top keywords, and snippets of example tweets.

Topics and top 20 keywords	Examples
Incidences of dengue (15.58 %): fever, die, dengue, doctor, kill, mosquito, flood, symptom, suffer, toll, treatment, woman, student, viral, road, cause, place, accident, bear, people	1. A woman dies of dengue fever at Dhaka Medical College Hospital #Dengue #Bangladesh. 2. #Dengue hits new heights due to floods. Vast number of deaths & sufferings are waiting for people.
Cases of hospitalization (17.47 %): dengue, patient, number, fever, admit, increase, hour, die, hospital, bed, people, year, case, country, suffer, outbreak, rise, day, record, death	1. Dengue cases are still surging across the country as 1,649 more new people were hospitalised with the mosquito-borne deadly virus in 48 hours #Bangladesh 2. More and more patients are now suffering from #dengue fever in Rangamati General Hospital.
Dengue diagnosis and treatment (15.65 %): dengue, test, hospital, free, treatment, fever, government, cost, clinic, patient, private, medical, kit, capital, bill, center, country, fee, diagnostic, case	1. Bangladesh Army to provide free dengue diagnosis and treatment. 2. Additional bills for dengue treatment across the country, the Consumer Rights Protection Department is coming to the grounds for inquiry.
Spread of dengue (21.51 %): dengue, mosquito, country, people, fever, Eid, control, body, water, prevent, spread, government, epidemic, emergency, panic, measure, disease, bite, spray, chemical	1. Dhaka dwellers beset by mosquitoes, rain poses risk of Dengue 2. #Beware of #Dengue mosquitoes from dawn till dusk ! Dengue carrying Mosquitoes usually bite on the legs in day time and have black and white stripes on their body. #Dengue #Epidemic #Bangladesh
Dengue prevention (17.92 %): dengue, prevent, awareness, mayor, campaign, fever, control, minister, program, rumor, mosquito, meeting, papaya, medicine, leaflet, bleach, rally, failure, initiative, clean	1. Dhaka Mayor has said conspirators are spreading rumors about dengue. 2. According to Director General of Health and quoted by News Media #dengue cases in #Bangladesh stand more than 24,500. @BDRCS1 with the support of @ifrc continues #awareness campaign all over the country through its volunteers network.
Blood donation (11.87 %): dengue, blood, need, patient, fever, emergency, bag, positive, ve, b, suffer, negative, old, brother, baby, year, save, help, urgent, critical	1. Urgent B(+) positive blood needed for a critical dengue patient at Dhaka. Cell: 01********* 2. My relative is suffering from dengue. Is there anyone who is willing to donate blood for her? We need AB+ blood donor who is currently in Dhaka.

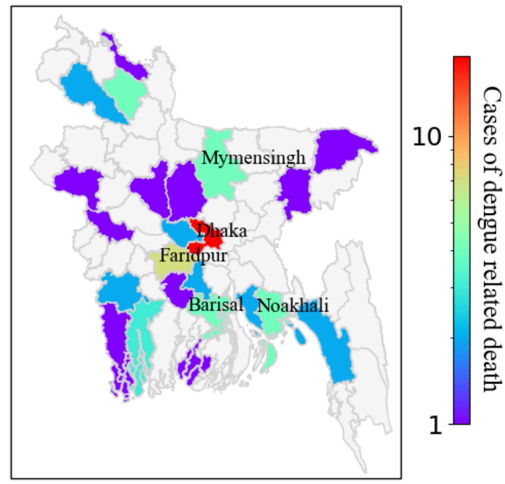
their concern in poor management of dengue, monsoon floods, road accidents, and other concurrent events that added to citizens' plight. One user tweeted:

> *"Outrageous situation for dengue epidemic, Aedes mosquitoes, flood, rumors, rape, road accidents, killings - Who is responsible for this? #Bangladesh #NoneToSee #BehindDevelopment"*

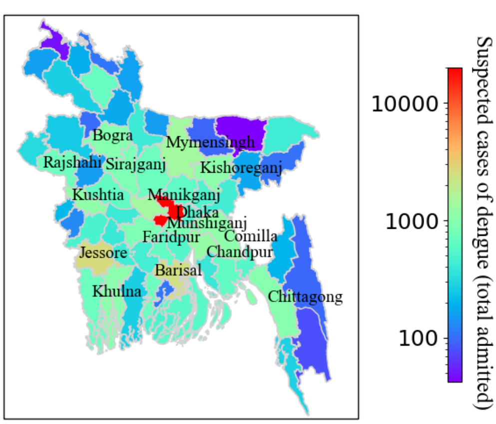
Apart from this, similar to another study in Brazil [73], we found that many people posted humorous tweets sharing their personal experiences with dengue and mosquitoes. For example, a user tweeted:



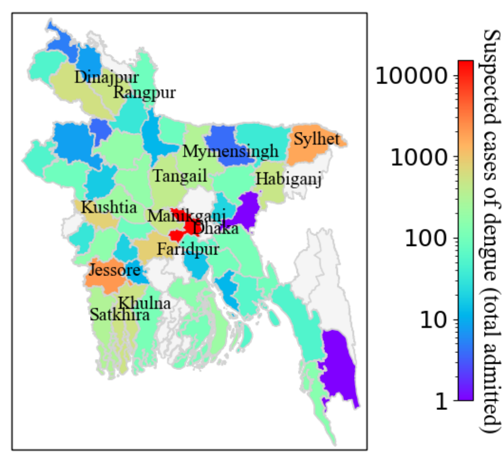
(a) Cases of death from government reports.



(b) Cases of death from tweets.



(c) Cases of hospitalization from government reports.



(d) Cases of hospitalization from tweets.

Fig. 3. District-wise distribution of dengue cases from January–September, 2019 as per government reports and tweets. Only the districts where the number of reported death cases and hospitalized dengue patients are greater than the third quartile (Q_3) of the corresponding cases are labeled with names. Color scales of the districts are presented in the logarithm of the reported cases.

> *"#Bengali_sentiment: O mosquito, you are the culprit! Give me back my blood. Who gave you the right to bite me? You won't be pardoned . . . "*

4.2 Cases of Hospitalization

About 18% tweets in our sample discussed the cases of hospitalization due to dengue. We first analyzed how much these tweets complement the data from government reports. DGHS health bulletins reported the total number of dengue patients hospitalized from January 2019–September 2019 for all 64 districts. On the contrary, dengue related tweets reported cases of hospitalization for 52 districts. A Wilcoxon signed-rank test showed that the reported cases of hospitalization in

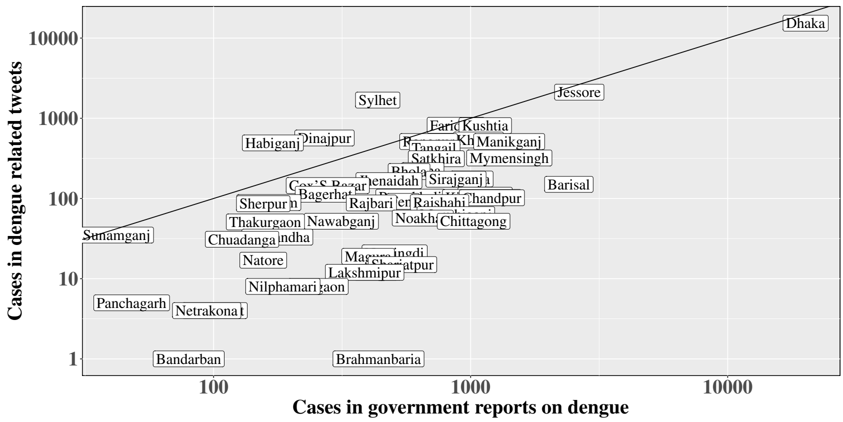


Fig. 4. District-wise anomaly in cases of hospitalization as per government reports and dengue related tweets. To facilitate visualization the cases are reported in log scale. Here, the straight line passes through the origin.

Table 4. Cases of dengue in different hospitals in Bangladesh from January–September, 2019.

Hospital type	Cases from government reports			Cases from tweets		
	N	Admitted (N)	Dead (N)	N	Admitted (N)	Dead (N)
Public	8	81	48	25	236	20
Private	18	40	84	6	33	0
Total	26	121	132	31	269	20

dengue related tweets significantly lag behind than that of the cases reported by the government ($W = 115.5$, $Z = -6.18$, $p < .001$, $r = 0.54$).

Figure 3c and 3d show spatial distribution of hospitalized dengue patients in different districts based on government reports and public tweets, respectively. Both datasets reported the highest number of hospitalized dengue patients in Dhaka ($N_{gov} = 20,100$ *and* $N_{Twitter} = 15,350$), followed by Jessore ($N_{gov} = 2,646$ *and* $N_{Twitter} = 2,117$). However, the ordering for the next three districts differed between the two datasets. While government data reported most cases in Barisal ($N = 2,411$), Mymensingh ($N = 1,414$), Manikganj ($N = 1,412$), Twitter data pointed to most cases in Sylhet ($N = 1,687$), Faridpur ($N = 818$), Kushtia ($N = 810$). Even though the ordering differed slightly, we found a medium size correlation between the cases of hospitalization reported by the government and by Twitter users ($r_\tau = 0.46$, $p < .01$).

To investigate further, we plotted the number of hospitalized dengue patients for each district in Figure 4. The districts that are close to the straight line or on it have little discrepancy in the Twitter- and government-reported cases of hospitalization (e.g., Dhaka, Jessore, Faridpur, Sunamganj). The districts that are below the line have underreported Twitter data (e.g., Bandarban, Brahmanbaria, Bairsal, Chittagong). The districts that are above the line (i.e., Habiganj, Dinajpur, and Sylhet) have less number of hospitalization cases reported by the government than by Twitter users. To investigate the reason behind this discrepancy, we manually examined 37 tweets that reported cases of hospitalization in Habiganj ($N = 7$), Dinajpur ($N = 18$), and Sylhet ($N = 12$). We found that 31 tweets were shared from the Twitter accounts of local or national news agencies and the remaining were citing these news reports, implying shortcomings in government reports. These findings indicate potential of mining Twitter data to address the gaps in government reports.

Next, we analyzed the cases of hospitalization and deaths in different hospitals in Bangladesh both from tweets and government reports (see Table 4). We observed that government reports covered public hospitals less than private hospitals. In contrast, Twitter users reported more about government hospitals than private hospitals. A Fisher's exact test indicated significant difference ($p < .001$) in the number of public and private hospitals mentioned across the two datasets. We also found that the number of deaths in public and private hospitals significantly differed ($t(27) = -2.2,\ p < .05$) across the two datasets. Compared to government reports, tweets reported significantly less deaths, perhaps due to poor availability and accessibility of government health reports in Bangladesh [114]. Although Bangladesh has implemented Right to Information Act to make it easier for public to access government data, the initiative is marred with dismal implementation [9].

With respect to dengue patients admitted to public hospitals, we found that the number reported by the government was much less than those reported in tweets. A Chi-square test indicated a significant effect of data source on the distribution of admitted patients in public and private hospitals ($\chi^2(1, N = 390) = 22.36,\ p < .001$). To examine further, we analyzed all the tweets ($N = 124$) that reported cases of hospitalization in different public hospitals. We found that 79 of these tweets were posted from the Twitter accounts of local or national news agencies and the remaining were citing these news reports, indicating potential gaps in government's account of dengue patients admitted to different public hospitals.

4.3 Dengue Diagnosis and Treatment

About 16% tweets in our sample focused on dengue diagnosis and treatment options in Bangladesh. Many tweets discussed schemes, programs, and resources available for speedy diagnosis of dengue. For example, many people shared how some hospitals, institutions, and Bangladesh army were providing free dengue tests. Others tweeted government's declaration of free treatment for dengue patients, free diagnosis at public hospitals, and capped costs of diagnosis at private hospitals.

> *"#Dhaka South #City Corporation (DSCC) Mayor #SayeedKhokon has said that if anyone in the #DSCC area is affected by #Dengue then please call our hotline number: 096********; #health workers will go to you and provide #care and #medicine free of cost."*

> *"#Special_post: if any dengue patient comes to either Uttara or Mirpur branch of R***** hospital, they will receive free test and no amount will be charged for the treatment.".*

> *"The fee of dengue test will be 500 tk. at private hospitals and free in public hospitals: Ministry of Health".*

However, despite different measures from the government, many tweets reported dismal implementation of these efforts, reflecting ground realities. People mentioned a lack of dengue test kits required for diagnosis. The shortage of kits greatly affected the patients and deprived many of proper diagnosis. Many tweets also reported a shortage of other healthcare facilities, such as, insufficient hospital beds to accommodate the huge influx of dengue patients. Even in some cases patients were denied treatment, creating panic and chaos. Some tweets complained how people were charged high fees for dengue testing, despite the declaration of free testing or capped costs. Moreover, some tweets reported about wrong diagnosis and ambiguous dengue test results, creating further confusion and worries among the patients.

> *"Test kit crisis hits dengue treatment in Dhaka as five more die in Bangladesh outbreak. Lack of kits for diagnosing dengue fever has hit the market with some hospitals in Dhaka turning the suspected patients away without testing them".*

> *"#Dengue test: K***** clinic gives different test #results for same blood."*

Table 5. Density of Aedes vectors at 14 public places in Dhaka as per DGHS survey [82].

Breteau Index				
<=20	20 - 40	40 - 60	60 - 80	80 +
Kalshi to ECB Chattar Bhashantek Slum	-	Metro Rail Project, Mirpur 12 Mohakhali Korail Slum Saydabad Bus Terminal Rajarbagh Police Lines	BRTC Bus Depot, Mirpur 12 Gabtoli Bus Terminal Dhaka Medical College Hospital	Mohakhali Bus Terminal Shahjahanpur Slum Mugda Medical College Railway Colony, Komlapur BRTC Bus Depot, Komlapur

To address this, the government fined several hospitals and diagnostic centers for charging dengue patients with high bills. We found several tweets mentioning government's action to enforce the policy.

> *"Taking extra fee for dengue tests: 4 clinics are fined 1.8 million tk in the capital"*

Our findings indicate Twitter's potential as a tool to assess ground realities and discover irregularities in the implementation of dengue outbreak response plan. For example, in our work, monitoring and analyzing these tweets could help policymakers identify issues that needed immediate attention, such as regulating costs of diagnosis and treatment, procuring diagnostic kits, opening emergency units in the hospitals with additional beds, and ensuring the quality of diagnosis.

4.4 Spread of Dengue

Nearly 22% tweets discussed the role of Aedes vector and human movements in the spread of dengue viral infection.

4.4.1 Aedes Vector and Dengue. Since Aedes mosquitoes play a key role in the transmission of dengue, we decided to examine the government reported distribution of Aedes vectors within the capital city of Dhaka [81, 82]. Table 5 shows the density of Aedes vectors in Breteau index[2](BI) for different wards[3] in Dhaka. Figure 5a and 5b show the spatial distribution of BI values and dengue cases for different areas, respectively.

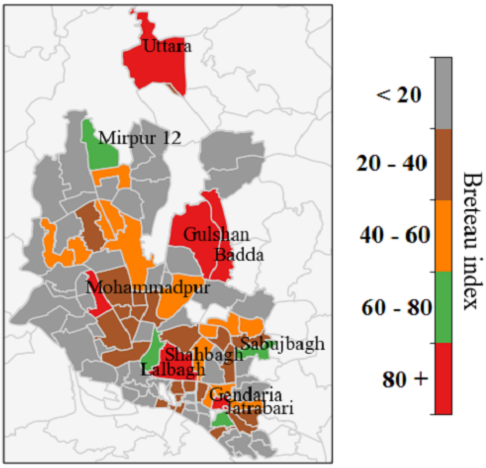
Our analysis revealed some correlation between the BI values reported by the government and the number of dengue cases based on tweets. For example, Mirpur had the highest dengue cases ($N = 73$) and places in it (e.g., BRTC bus depot, Mirpur 12) were classified as highly dense breeding places of Aedes vectors. Similarly, Uttara had the second highest dengue cases ($N = 43$) and a high BI score (80+) score. We found similar trends for many other areas, such as Mohammadpur ($N = 12$) and Gulshan ($N = 11$), indicating usefulness of tweets to monitor local transmission of dengue as government health bulletins often do not report area-wise cases of dengue. Even media reports corroborated our findings that these local areas had higher number of dengue cases [10].

4.4.2 Eid-ul-Adha and Dengue. Many tweets expressed concern that during Eid-ul-Adha, one of the main religious festivals of the Muslims, dengue infection would spread across the country because of large human movements. Therefore, some tweets requested people not to travel to their village homes during Eid.

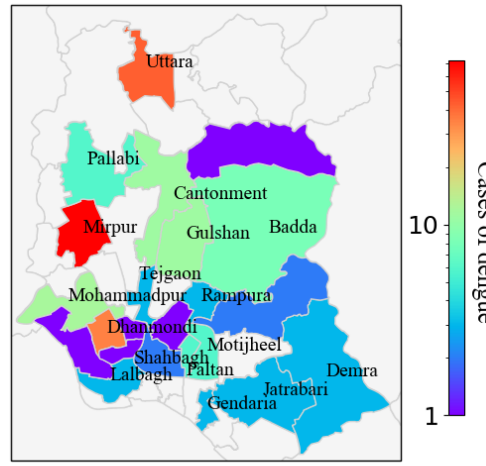
> *"Qurbani Eid holiday is close, people will leave Dhaka and go to the village, dengue patients will also travel. If a mosquito bites a patient, there are chances that the disease will spread through that mosquito. If you care for your loved ones, do not travel."*

We decided to check whether these concerns were justified. We first analyzed the temporal distribution of dengue cases both from the government reports and the dengue related tweets.

[2]Breteau index is the number of Aedes positive containers per 100 houses inspected [113]
[3]A ward is low-level administrative division of a city.



(a) Distribution of Aedes vectors within Dhaka.



(b) Distribution of dengue cases within Dhaka.

Fig. 5. Figure 5a shows DGHS reported spatial distribution of Aedes vectors in different areas of Dhaka. Only the names of areas with Breteau Index (BI)>= 60 are displayed. Figure 5b shows the cases of dengue in different areas of Dhaka from the tweets. Only the names of areas where the number of dengue cases is greater than the third quartile (Q_3) of the cases in all areas are displayed. Color scales of the areas are presented in the logarithm of the reported dengue cases.

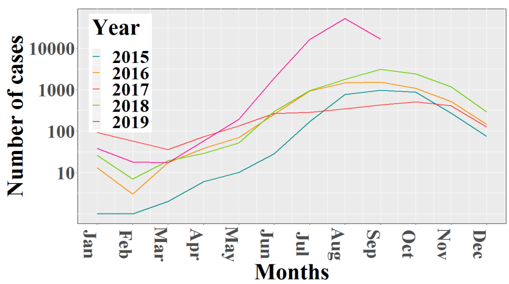
Table 6. Total number of people infected/ hospitalized/ dead due to dengue during Eid-ul-Adha (from tweets).

Year	2015	2016	2017	2018	2019
One week (Before, After)	(0, 3)	–	(0, 1200)	(1, 1034)	(9050, 62942)
Two weeks (Before, After)	(2, 2300)	–	(0, 1201)	(1, 3799)	(24822, 69533)

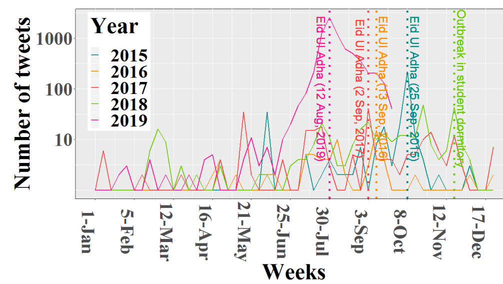
A positive correlation ($r_\tau = 0.51$, $p < .001$) between the government reported dengue cases and the frequency of dengue related tweets per month suggests that people engage more in disease related conversations as the intensity of dengue rises. Figure 6a and 6b plot the number of monthly dengue cases based on government reports and the weekly distribution of dengue related tweets, respectively. From IEDCR reports, we can see that the severity of dengue is higher from August–September (Figure 6a), which is the post monsoon period and the time for celebrating Eid-ul-Adha in Bangladesh [77]. Even the weekly distribution of dengue related tweets in a given year reaches its peak around Eid-ul-Adha (Figure 6b).

To further examine, we compared the user-reported cases of dengue both within two weeks and within one week interval of Eid-ul-Adha from 2015–2019 (see Table 6). We found that more people were affected by dengue after Eid-ul-Adha every year except in 2016 where tweets in our sample reported only qualitative information, such as *"number of dengue victims rise in Dhaka."*

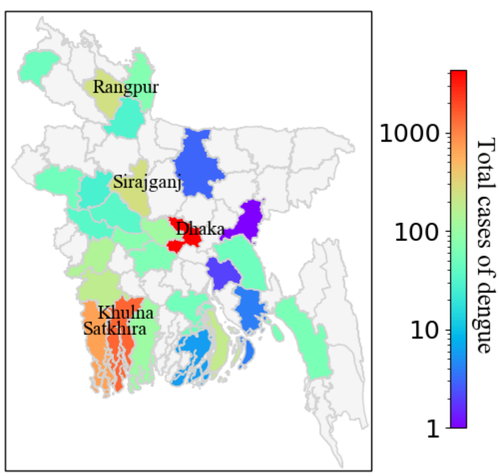
We then examined spatial distribution of the dengue cases before and after Eid-ul-Adha. For this analysis, we only considered tweets posted in 2019 due to sparse availability of tweets in other years. Figure 6c and 6d plot the spatial distribution of dengue cases within one week interval of Eid-ul-Adha. A visual inspection of these figures indicate that dengue infection spread to 7 new districts.



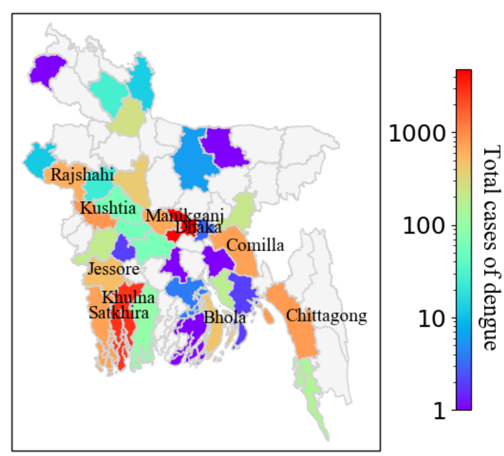
(a) Government reported cases of dengue per month.



(b) Weekly frequency of dengue related tweets.



(c) District-wise distribution of dengue cases one week before Eid-ul-Adha.



(d) District-wise distribution of dengue cases one week after Eid-ul-Adha.

Fig. 6. Frequency of dengue cases in Bangladesh. Figure 6a shows IEDCR reported monthly distribution of dengue cases from 2015–2019. Figure 6b shows the weekly distribution of tweets regarding dengue outbreaks in Bangladesh from 2015–2019. To enable better visualization, we plot the numbers in Figure 6a and 6b in log scale. Figure 6c and 6d show district-wise distribution of dengue cases (total number of people infected, hospitalized, or dead due to dengue) in one week interval of 2019's Eid-ul-Adha in Bangladesh from the tweets. Color scales of the districts are presented in the logarithm of the reported dengue cases.

A Wilcoxon signed-rank test revealed that the number of dengue patients was significantly greater ($W = 715.5$, $Z = 2.98$, $p < .01$, $r = 0.36$) a week after the Eid than the week before, suggesting that the government should put precautionary measures in place during Eid to cope with sudden outbreaks in new areas.

Note that in 2018, the number of dengue cases reported on Twitter also soared in winter (see Figure 6b). This is because many university students living in dormitories got dengue infection, leading to an outbreak.

> *"Many from Dhaka University, student dormitories, mess halls, and sublet homes are affected by dengue. Shouldn't we close the educational institutes under this circumstances?"*

Government surveys on dengue are usually limited to a number of public places, commercial, or residential areas. However, these tweets indicate the need to survey dormitories and other places where many people either gather or live in close contact.

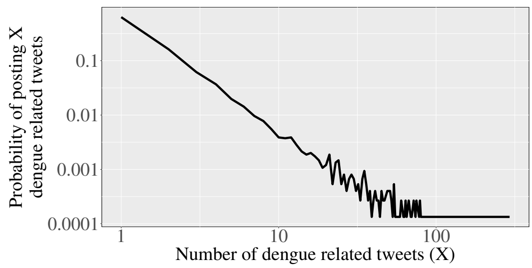


Fig. 7. Probability distribution of the number of dengue related tweets, per Twitter user. The plot has been generated in log-log scale to improve visualization.

4.5 Dengue Prevention

About 18% tweets discussed dengue awareness and prevention programs as well as government's failure in controlling the epidemic. Many tweets discussed how different organizations took measures, such as spraying insecticides to control Aedes mosquitoes, raising public awareness by distributing leaflets, and organizing awareness rally and cleanliness program, to prevent the spread.

> *"Today @BDRCS1 RCY #Volunteers in Cox's Bazar organized a rally in the city, distributing leaflets to raise awareness on #Dengue #outbreak . This is part of the Bangladesh Red Crescent response to prevent further spread of the infections"*

> *"North City Corporation collects machinery and pesticide to prevent dengue"*

Next, we examined Twitter accounts which were most active in disseminating dengue related information. To do this, we calculated the number of dengue related tweets posted by each Twitter account in our sample. Figure 7 shows the probability distribution of tweeting about dengue. In the figure, x-axis represents the number of tweets about dengue (X) and y-axis represents the probability that users will post X number of dengue related tweets. As expected, a majority of users (63%, $N = 4,736$) posted only one tweet about dengue. In contrast, top 1%, top 10%, and top 20% users were responsible for 25%, 58%, and 69% of tweets, respectively. The most active user (jugjugantor24, an online news platform) posted 293 dengue related tweets.

On average, users posted 3 dengue related tweets. Among the $1,070$ users who tweeted about dengue more than the average, 566 (52.9%) users are either news media, international organizations, journalists, or doctors. These 566 users posted over 10,000 tweets collectively (average = 19 tweets). We also examined how other users interacted with dengue related tweets. On average, these tweets were retweeted 9 times. Among the 126 users whose posts were retweeted more than the average, 82 (65%) are Twitter accounts of news and media agencies. On average, the posts of these 82 users were retweeted 46 times. Table 7 shows the top users with most dengue related tweets and retweets. User1, User2, and User3 are a social entrepreneur, coordinator of an online blood donation group, and a student political leader, respectively, and the rest are accounts of news and media agencies.

We found that the government institutions were almost inactive on Twitter during the epidemic; ICDDR[4] posted 19 tweets, WHO's regional office posted 1 tweet, and DGHS posted no tweets, indicating lack of efforts by these agencies in using Twitter to communicate disease related information to the public. The shortcomings in government's dengue outbreak response plan and a lack of direct communication created a sense of dissatisfaction among the public.

Despite heavy criticism on social media, the authorities constantly denied all the complaints saying that the chaos around dengue was nothing but *rumor*. For example, the Mayor of the capital

[4]ICDDR is International Centre for Diarrhoeal Disease Research in Bangladesh, https://www.icddrb.org/

Table 7. Some of the top Twitter users who tweeted most frequently about dengue and whose posts were retweeted most during epidemic. Personal accounts are replaced with User to preserve anonymity.

Top users who tweeted about dengue most frequently	Number of dengue related tweets	Users with highest number of retweets	Number of retweets
jugjugantor24	293	dailystarnews	509
dailystarnews	194	bbcbangla	440
User1	178	User3	164
barta24bd	169	AmraTbashi	146
TheDailyInqilab	159	samakaltw	141
User2	154	AFP	124
DhakaTimes	148	ProthomAlo	118
samakaltw	133	AITCofficial	117
Nagorik2	127	RSTMH	105
banglanews_eng	123	DAILYITTEFAQ	104

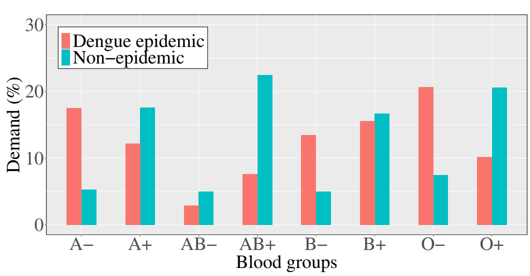


Fig. 8. Demand for blood groups in Bangladesh during dengue epidemic and non-epidemic period from personal blood donation tweets.

Table 8. Distribution of blood groups within Bangladeshi population obtained using [34] and blood donors obtained using [2].

Blood group	Original population (%)	Blood donors (%)
A-	0.67	0.58
A+	26.01	23.59
AB-	0.27	0.29
AB+	8.29	7.98
B-	0.7	5
B+	33.66	33.54
O-	0.96	1.03
O+	29.44	32.02

city Dhaka claimed that the news about the rise of dengue was another rumor. We also found some tweets that informed the public of myths around dengue prevention. For example, a rumor spread that bleaching powder could prevent dengue. A user tweeted the following to dispel the rumor:

> *"#Be aware of new rumor. A new rumor is being circulated on Facebook and WhatsApp that by pouring bleaching powder and harpic together in toilets, dengue can be prevented."*

We discuss misinformation and rumors in more detail in Section 5.

4.6 Blood Donation

Around 12% tweets were about blood donation for dengue patients. Figure 8 shows the demand for different blood groups during dengue epidemic based on personal blood donation requests on Twitter. The highest demands were for O- (20.63%), A- (17.47%), and B+ (15.57%) blood groups. In addition, some tweets also mentioned an urgent need of blood transfusion for patients with very low ($< 10,000$) platelet counts. To further examine, we compared the demand for blood transfusions during dengue epidemic with the demand during the non-epidemic period.

Since distribution of blood group varies with ethnicity, we first gathered data from prior studies to understand the distribution of blood groups [34] and of blood donors in Bangladesh [2]. We found a significant positive correlation ($r = 0.99$, $p < .001$) between the two distributions. Table 8 shows that Rh positive blood groups (i.e., B+, O+, A+, AB+) are much more dominant in Bangladesh compared to Rh negative blood groups. Our analysis of the tweets during non-epidemic period indicated a variety of health conditions, such as kidney transplant, heart surgery, cancer, tumor operation, and childbirth, for which people requested blood donations.

Figure 8 plots the demand for different blood groups on Twitter during the epidemic and non-epidemic period. Overall, we found more demand for blood transfusion during the epidemic period. A Wilcoxon signed-rank test revealed significant differences in the demand for blood transfusion ($W = 18,188$, $Z = 3.37$, $p < .001$, $r = 0.39$) during dengue epidemic (3.7 bags on average, $SD = 3.9$ bags) compared to the non-epidemic period (2.5 bags on average, $SD = 2.4$ bags). We also found that the distribution of demand for different blood groups varied significantly ($\chi^2(7) = 27.949$, $p < .001$) during dengue epidemic from the demand during non-epidemic period. For example, we found that A-, B-, and O- blood groups were in more demand than other blood groups during the epidemic period. Perhaps the scarcity of Rh negative blood donors led to high demand for these blood groups on Twitter during dengue epidemic. We found little fluctuations in the comparative demands of B+ and AB- blood groups, probably because B+ and AB- are the most and least abundant blood groups in Bangladesh, respectively. Due to the abundance of B+ blood donors (33.54%), people may hardly face any difficulties in finding a donor either through personal contacts or other means. On the other hand, due to the scarcity of AB- blood group among Bangladeshi population (0.27%), its distribution is supposed to be less among dengue affected patients, leading to less demand on Twitter. Also, the differences between the proportion of people with B+ and AB- blood groups and available blood donors for them is the least compared to other blood groups as shown in Table 8.

Overall, the analysis of blood donation requests on Twitter showed that the demand for different blood groups vary significantly during dengue epidemic, suggesting the need to take proactive measures to prevent serious shortages of blood supply and the subsequent loss of lives.

5 DISCUSSION

In this work, we conducted a longitudinal analysis of tweets to identify public health issues that emerge during dengue outbreaks in Bangladesh. We found that people used Twitter to access and share dengue related information, report mismanagement in dengue related healthcare services, dispel popular myths and rumors, and request resources from public (e.g., blood donation requests) to fight the infection. On the other hand, we found that the government healthcare bodies were largely inactive on Twitter and did little to spread awareness and disseminate information about outbreaks online. We also examined how insights from tweets complemented the data present in obscure government reports and found several examples where government account of dengue infections and deaths were lacking. Our analysis also revealed novel insights about social, cultural, geographical, and epidemiological aspects of dengue epidemic in Bangladesh. For example, we found how human movements during Eid-al-Adha resulted in peak infections and new outbreaks, how women were dying more from dengue infection, and how new patterns of blood donation requests emerged during dengue epidemic. Based on our findings from dengue related tweets about the healthcare system in Bangladesh (RQ1), we now discuss how the insights obtained from tweets can inform policy making approaches during emergency health crisis such as dengue epidemic (RQ2).

5.1 Disease Surveillance

Our analysis indicated shortcomings in dengue statistics compiled by government agencies. For example, Twitter dataset indicated more deaths in 13 districts and more cases of hospitalization in three districts. We also found a discrepancy in the estimates from international health organizations and the cases reported by the government. For example, WHO estimated over 350,000 dengue infections in July 2019 while the government reported only 7,179 official cases [57]. Improper statistics for infectious diseases like dengue can paralyze outbreak response plan of government as well as non-government agencies, leading to a mismanagement of already limited resources and loss of human lives.

While by no means Twitter can be used for discovering the exact number of dengue cases, it can be a powerful tool to reveal discrepancy in the reporting of dengue infections and deaths. For example, there are often undetected or unknown cases because of passive surveillance and inappropriate methods adopted by the government [57]. When data from tweets report unknown cases, deaths, or outbreak locations, government agencies can ask local and regional authorities for verification, springing them into action. Regional, national, and international healthcare organizations responding to dengue outbreaks can re-allocate resources to areas where new infections and outbreaks are found. Similar to [61, 67, 80], government agencies could also consider integrating social media health data into disease surveillance programs.

Practical challenges like tackling misinformation and analyzing code-mixed tweets can also make Twitter an unwieldy tool. Even in our dataset, we encountered tweets spreading rumors and myths related to dengue diagnosis and treatment. Social media platforms, such as Twitter and Facebook face grand challenges in discovering and controlling misinformation, which can particularly prove fatal during public health crises [85]. For example, rumors about COVID-19 on social media have increased panic and racist attitudes among people [12]. Also, verifying information in each tweet is a time and resource intensive undertaking. One way to address the misinformation challenge is to rely more on trusted sources like reputed news and media organizations, international health organizations, and non-governmental organizations with significant on-the-ground presence. To dispel rumors and myths related to dengue, government agencies may consider adopting social media to share information and updates about the latest developments of epidemic.

Another challenge in extracting insights from tweets is code-switching. Although we only extracted tweets in English by using TweetScraper, we found nearly 400 tweets where Bengali was intermixed with English in a sentence. Current NLP models are often inadequate for low-resource languages like Bengali, demand prior information on code-mixed languages, and require a sizeable corpus of code-mixed tweets [94]. Although recent NLP advancements can identify English and Bengali words in code-mixed sentences [13, 23], further research is needed to automate real-time analysis of code-switched tweets to extract meaningful insights.

5.2 Controlling Local Transmission of Diseases

Apart from increasing government's accountability and transparency, tweets on public health issues can also help government agencies and healthcare institutions control local transmission of diseases. For example, many tweets on dengue mentioned areas with high infection rates in the capital city Dhaka. The government may leverage this bottom-up, citizen-reported, decentralized information to plan effective intervention programs for controlling dengue instead of spraying invasive chemicals indiscriminately in all areas. For example, our analysis of tweets suggested that government should use pesticides in water-clogged areas such as Mirpur, but may adopt non-invasive intervention programs (e.g., awareness and cleanliness drives) for areas with low Aedes density, such as Dhanmondi. The government can also use tweets as a source of information to detect early-stage dengue infections at unexpected locations. For example, some tweets indicated local transmission of dengue in the student dormitories, a private area that is rarely checked for the presence of Aedes vectors by the administration. Timely access to such information can help the government take necessary measures, such as social distancing [37], to prevent outbreaks in places where many people live together.

Government agencies and healthcare institutions can also analyze historical tweets to identify recurring events that contribute to dengue outbreaks. For example, we found that large human movements during Eid-ul-Adha resulted in unprecedented number of dengue infections and new outbreaks. While it is difficult to limit people's movement during a religious festival like Eid due to socio-cultural norms, the government can enhance cleanliness measures in public places with high footfall, such as railway stations, bus terminals, and cattle markets set up during Eid, that are known to have a large number of breeding sites [5, 82]. The government can also take proactive measures, such as preparing hospitals in different districts beforehand, to cope with the sudden outbreaks of dengue.

Our spatio-temporal analysis of tweets identified how dengue infections were emerging in different geographic locations. Such real-time localized information when paired with other datasets (e.g., population density, historical cases of dengue infections, available testing and treatment facilities) can help government agencies identify areas that are under imminent risk, requiring immediate attention. This can also help the government prioritize the allotment of its limited resources to different areas based on the severity of dengue infections. Such strategy can be more effective than distributing resources evenly which may leave many areas under high risk with low budget.

Government agencies can also analyze tweets to discover other factors that impact their ability to control dengue outbreaks. For example, many tweets in our sample pointed to the co-occurrence of flood and dengue in Bangladesh. It has been reported that floods can potentially increase the transmission of dengue hemorrhagic fever [112]. Moreover, findings from previous study show that high river levels are strongly associated with increased cases of dengue fever in Dhaka [46]. Identifying such factors can help the government to organize a better response to dengue epidemic. For example, in case of flood, the government should focus first on the low-lying areas that are usually affected by floods while planning to control dengue.

5.3 Maintaining Healthcare Services

Apart from monitoring and controlling dengue infections, maintaining a functional healthcare system is equally important during an epidemic when a country's health care systems are overwhelmed with huge influx of patients. Personal experiences shared by people on Twitter can help the government and policymakers reflect on the taken measures and get a grasp of on-the-ground realities. For example, many tweets in our dataset reported shortages of dengue diagnostic kits, lack of access to hospital beds, and high costs of dengue diagnosis and treatment. More importantly, access to pain points and irregularities in implementation could help international health organizations and local non-governmental organizations streamline their efforts and resources.

In addition to streamlining reactive measures, government agencies and other health care organizations can analyze historical tweets to design informed proactive measures. A compelling example is blood donation requests on Twitter. Often high-risk dengue patients require blood and platelet transfusion to prevent hemorrhagic complications. To date, there is a dearth of research on how people use social media to look for volunteer blood donors. Our analysis of personal blood donation requests on Twitter showed that the Rh negative blood groups were in high demand during dengue epidemic compared to the non-epidemic period, suggesting that hospitals should stock Rh negative blood groups more in preparation of dengue outbreaks. Such insights can minimize the loss of lives by helping healthcare professionals be well-prepared in the event of an outbreak. Future research could explore what motivates people to request blood donations on Twitter, how donors respond to such requests, how quickly requesters find a donor, and how network and messaging could improve yield.

6 LIMITATIONS

Our work has some limitations. Firstly, for our analysis, we only considered the tweets composed in English because of the challenges in using current NLP models for identification, translation, and topic modeling of Bengali. We recognize that by doing so, we miss out on the nuances and contexts present in tweets written in Bengali. Secondly, although misinformation and rumors are often present in social media data [38, 97], we did not verify information reported in tweets except when our analysis indicated shortcomings in government reports. Further work is needed to identify and tackle misinformation in tweets posted during public health crises like dengue epidemic. Thirdly, for our analysis, we relied only on Twitter data and could not analyze Facebook posts. Even though Facebook has the largest user base in Bangladesh among all social media platforms [105], we could not utilize it due to limited access to Facebook's Public Feed API [110]. Future research could use other sources like Facebook and WhatsApp to triangulate insights obtained from analyzing social media health data.

7 CONCLUSION

We present an in-depth analysis of the tweets related to dengue epidemic in Bangladesh to explore the role of social media health data in informing policy making approaches. We apply a human-centered lens to discover insights from the reactions and responses of Twitter users to dengue epidemic in Bangladesh. Our analysis identified various epidemiological, geographical, cultural, and societal factors that mediated the outcomes of dengue epidemic in Bangladesh, which were not perceptible at the first glance. We also showed how our findings could inform policy making approaches to control dengue outbreaks. We believe this work will greatly facilitate social media supported cooperative works to cope with public health crises in developing countries.

REFERENCES

[1] Rabeeh Ayaz Abbasi, Onaiza Maqbool, Mubashar Mushtaq, Naif R. Aljohani, Ali Daud, Jalal S. Alowibdi, and Basit Shahzad. 2018. Saving lives using social media: Analysis of the role of twitter for personal blood donation requests and dissemination. *Telematics and Informatics* 35, 4 (2018), 892 – 912. https://doi.org/10.1016/j.tele.2017.01.010

[2] M A Ahad, M A Bakar, and H A M Nazmul Hasan. 2009. Pattern of ABO and Rhesus (Rh) Blood Group Among Blood Donors. *TAJ: Journal of Teachers Association* 15, 2 (2009), 68–70. https://doi.org/10.3329/taj.v15i2.3910

[3] Hussein Al-Olimat, Krishnaprasad Thirunarayan, Valerie Shalin, and Amit Sheth. 2018. Location Name Extraction from Targeted Text Streams using Gazetteer-based Statistical Language Models. In *Proceedings of the 27th International Conference on Computational Linguistics* (Santa Fe, New Mexico, USA). Association for Computational Linguistics, Stroudsburg, Pennsylvania, USA, 1986–1997. https://www.aclweb.org/anthology/C18-1169

[4] Hussein S. Al-Olimat. 2018. LNEx: Location Name Extractor. https://github.com/halolimat/LNEx

[5] Helemul Alam and Mahbubur Rahman Khan. 2019. Dengue Outbreak: Cattle markets may heighten risks. Retrieved March 27, 2020 from https://www.thedailystar.net/frontpage/news/dengue-outbreak-cattle-markets-heighten-risks-1783894

[6] Turki Alanzi and Batool Alsaeed. 2019. Use of Social Media in the Blood Donation Process in Saudi Arabia. *Journal of Blood Medicine* 10, null (2019), 417–423. https://doi.org/10.2147/JBM.S217950

[7] Julio Albinati, Wagner Meira, Gisele L. Pappa, Mauro Teixeira, and Cecilia Marques-Toledo. 2017. Enhancement of Epidemiological Models for Dengue Fever Based on Twitter Data. In *Proceedings of the 2017 International Conference on Digital Health* (London, United Kingdom) *(DH '17)*. Association for Computing Machinery, New York, NY, USA, 109–118. https://doi.org/10.1145/3079452.3079464

[8] Ali Alessa and Miad Faezipour. 2018. A review of influenza detection and prediction through social networking sites. *Theoretical Biology and Medical Modelling* 15, 2 (2018), 27. https://doi.org/10.1186/s12976-017-0074-5

[9] Abdullah Alif. 2019. 10 years of Right to Information Act: 59% of appellants get no response from authorities. Retrieved March 26, 2020 from https://www.dhakatribune.com/bangladesh/2019/10/29/10-years-of-right-to-information-act-59-of-appellants-get-no-response-from-authorities

[10] Prothom Alo. 2019. How dwellers of Dhanmondi, Mohammadpur suffer from dengue. Retrieved March 26, 2020 from https://en.prothomalo.com/bangladesh/How-dwellers-of-Dhanmondi-Mohammadpur-suffer-from

[11] Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, and Joseph Bayer. 2016. Developing a Research Agenda for Human-Centered Data Science. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (San Francisco, California, USA) *(CSCW '16 Companion)*. Association for Computing Machinery, New York, NY, USA, 529–535. https://doi.org/10.1145/2818052.2855518

[12] Billy Bambrough. 2020. Coronavirus COVID-19 Will Go Down In History As The Social Media And Bitcoin Pandemic. Retrieved March 26, 2020 from https://www.forbes.com/sites/billybambrough/2020/03/23/coronavirus-covid-19-will-go-down-in-history-as-the-social-media-and-bitcoin-pandemic/#7c6d400b3c10

[13] Utsab Barman, Amitava Das, Joachim Wagner, and Jennifer Foster. 2014. Code Mixing: A Challenge for Language Identification in the Language of Social Media. In *Proceedings of The 1st Workshop on Computational Approaches to Code Switching* (Doha, Qatar). Association for Computational Linguistics, Stroudsburg, PA, USA, 13–23.

[14] Asha A. Bharambe and Dhananjay R. Kalbande. 2016. Techniques and Approaches for Disease Outbreak Prediction: A Survey. In *Proceedings of the ACM Symposium on Women in Research 2016* (Indore, India) *(WIR '16)*. Association for Computing Machinery, New York, NY, USA, 100–102. https://doi.org/10.1145/2909067.2909085

[15] Stuart J. Blair, Yaxin Bi, and Maurice D. Mulvenna. 2020. Aggregated topic models for increasing social media topic coherence. *Applied Intelligence* 50, null (2020), 138–156. https://doi.org/10.1007/s10489-019-01438-z

[16] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, null (2003), 993–1022.

[17] Todd Bodnar and Marcel Salathé. 2013. Validating Models for Disease Detection Using Twitter. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13 Companion)*. Association for Computing Machinery, New York, NY, USA, 699–702. https://doi.org/10.1145/2487788.2488027

[18] E Botelho-Nevers and P Gautret. 2013. Outbreaks associated to large open air festivals, including music festivals, 1980 to 2012. *Eurosurveillance* 18, 11 (2013), 9. https://doi.org/10.2807/ese.18.11.20426-en

[19] David A. Broniatowski, Michael J. Paul, and Mark Dredze. 2013. National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic. *PLOS ONE* 8, 12 (12 2013), e83672. https://doi.org/10.1371/journal.pone.0083672

[20] Caroline O. Buckee, Amy Wesolowski, Nathan N. Eagle, Elsa Hansen, and Robert W. Snow. 2013. Mobile phones and malaria: Modeling human and parasite travel. *Travel Medicine and Infectious Disease* 11, 1 (2013), 15 – 22. https://doi.org/10.1016/j.tmaid.2012.12.003 Special Issue : Malaria Reviews.

[21] Cody L. Buntain and Jung Kyu Rhys Lim. 2018. #pray4victims: Consistencies in Response to Disaster on Twitter. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 18. https://doi.org/10.1145/3274294

[22] Emily H. Chan, Vikram Sahai, Corrie Conrad, and John S. Brownstein. 2011. Using Web Search Query Data to Monitor Dengue Epidemics: A New Model for Neglected Tropical Disease Surveillance. *PLOS Neglected Tropical Diseases* 5, 5 (2011), 1–6. https://doi.org/10.1371/journal.pntd.0001206

[23] Arunavha Chanda, Dipankar Das, and Chandan Mazumdar. 2016. Unraveling the English-Bengali Code-Mixing Phenomenon. In *Proceedings of The 2nd Workshop on Computational Approaches to Code Switching* (Austin, Texas). Association for Computational Linguistics, Stroudsburg, PA, USA, 80–89.

[24] Andreas Charalambous. 2019. Social Media and Health Policy. *Asia-Pacific Journal of Oncology Nursing* 6, 1 (2019), 24–27. https://doi.org/10.4103/apjon.apjon_60_18

[25] Lauren E. Charles-Smith, Tera L. Reynolds, Mark A. Cameron, Mike Conway, Eric H. Y. Lau, Jennifer M. Olsen, Julie A. Pavlin, Mika Shigematsu, Laura C. Streichert, Katie J. Suda, and Courtney D. Corley. 2015. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. *PLOS ONE* 10, 10 (2015), 1–20. https://doi.org/10.1371/journal.pone.0139701

[26] Cynthia Chew and Gunther Eysenbach. 2010. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS One* 5, 11 (Nov. 2010), e14118. https://doi.org/10.1371/journal.pone.0014118

[27] Rumi Chunara, Jason R. Andrews, and John S. Brownstein. 2012. Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak. *The American Journal of Tropical Medicine and Hygiene* 86, 1 (2012), 39–45. https://doi.org/10.4269/ajtmh.2012.11-0597

[28] Jacqueline S. Coberly, Clayton Fink, Yevgeniy Elbert, In-Kyu Yoon, John Mark S Velasco, Agnes D. Tomayao, Enrique A. Tayag, Durinda R. Macasocol, and Sheri Lewis. 2014. Tweeting Fever: Can Twitter Be Used to Monitor the Incidence of Dengue-Like Illness in the Philippines? *JOHNS HOPKINS APL TECHNICAL DIGEST* 34, 4 (2014), 714–725.

[29] Bruna de Paula Fonseca e Fonseca and Fabio Zicker. 2016. Dengue research networks: building evidence for policy and planning in Brazil. *Health Research Policy and Systems* 14, 80 (2016), 10. https://doi.org/10.1186/s12961-016-0151-y

[30] Angel N. Desai, Moritz U. G. Kraemer, Sangeeta Bhatia, Anne Cori, Pierre Nouvellet, Mark Herringer, Emily L. Cohn, Malwina Carrion, John S. Brownstein, Lawrence C. Madoff, and Britta Lassmann. 2019. Real-time Epidemic Forecasting: Challenges and Opportunities. *Health Security* 17, 4 (2019), 268–275. https://doi.org/10.1089/hs.2019.0022

[31] DGHS. 2019. Daily Dengue Status Report. Retrieved October 23, 2019 from https://dghs.gov.bd/index.php/bd/home/5200-daily-dengue-status-report

[32] DGHS. 2019. Division and District wise Institutes List. Retrieved January 4, 2020 from http://app.dghs.gov.bd/inst_info/dist_list.php

[33] Parnali Dhar-Chowdhury, Kishor Kumar Paul, C. Emdad Haque, Shakhawat Hossain, L. Robbin Lindsay, Antonia Dibernardo, W. Abdullah Brooks, and Michael A. Drebot. 2017. Dengue seroprevalence, seroconversion and risk factors in Dhaka, Bangladesh. *PLOS Neglected Tropical Diseases* 11 (2017), 1–24. 10.1371/journal.pntd.0005475

[34] Tashmim Farhana Dipta, Md. Roushan Iqbal, Ahmed Zahid Hossain, Md. Tahminur Rahman, and Subhagata Chowdhury. 2011. Distribution of Phenotypic and Genotypic ABO and Rhesus Blood Groups Among Bangladeshi Population. *Ibrahim Medical College Journal* 5, 2 (2011), 59–62. https://doi.org/10.3329/imcj.v5i2.10101

[35] Luis Fernández-Luque and Teresa Bau. 2015. Health and Social Media: Perfect Storm of Information. *Healthcare Informatics Research* 21, 2 (2015), 67–73. https://doi.org/10.4258/hir.2015.21.2.67

[36] European Centre for Disease Prevention and Control. 2019. Dengue worldwide overview. Retrieved January 2, 2019 from https://www.ecdc.europa.eu/en/dengue-monthly

[37] European Centre for Disease Prevention and Control. 2020. Considerations relating to social distancing measures in response to COVID-19 – second update. Retrieved March 27, 2020 from https://www.ecdc.europa.eu/en/publications-data/considerations-relating-social-distancing-measures-response-covid-19-second

[38] Isaac Chun-Hai Fung, Carmen Hope Duke, Kathryn Cameron Finch, Kassandra Renee Snook, Pei-Ling Tseng, Ana Cristina Hernandez, Manoj Gambhir, King-Wa Fu, and Zion Tsz Ho Tse. 2016. Ebola virus disease and social media: A systematic review. *American Journal of Infection Control* 44, 12 (2016), 1660 – 1671. https://doi.org/10.1016/j.ajic.2016.05.011

[39] Sebastian Funk, Anton Camacho, Adam J. Kucharski, Rachel Lowe, Rosalind M. Eggo, and W. John Edmunds. 2019. Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15. *PLOS Computational Biology* 15 (2019), 1–17. https://doi.org/10.1371/journal.pcbi.1006785

[40] Philippe Gautret and Robert Steffen. 2016. Communicable diseases as health risks at mass gatherings other than Hajj: what is the evidence? *International Journal of Infectious Diseases* 47 (2016), 46 – 52. https://doi.org/10.1016/j.ijid.2016.03.007 Mass Gathering Medicine.

[41] Shalmoli Ghosh, Koustav Rudra, Saptarshi Ghosh, Niloy Ganguly, Sanjay Podder, Naveen Balani, and Neville Dubash. 2019. Identifying Multi-Dimensional Information from Microblogs During Epidemics. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data* (Kolkata, India) *(CoDS-COMAD '19)*. Association for Computing Machinery, New York, NY, USA, 224–230. https://doi.org/10.1145/3297001.3297030

[42] Elizabeth M. Glowacki, Allison J. Lazard, Gary B. Wilcox, Michael Mackert, and Jay M. Bernhardt. 2016. Identifying the public's concerns and the Centers for Disease Control and Prevention's reactions during a health crisis: An analysis of a Zika live Twitter chat. *American Journal of Infection Control* 44, 12 (Dec. 2016), 1709–1711. https://doi.org/10.1016/j.ajic.2016.05.025

[43] Janaína Gomide, Adriano Veloso, Wagner Meira, Virgílio Almeida, Fabrício Benevenuto, Fernanda Ferraz, and Mauro Teixeira. 2011. Dengue Surveillance Based on a Computational Model of Spatio-Temporal Locality of Twitter. In *Proceedings of the 3rd International Web Science Conference* (Koblenz, Germany) *(WebSci '11)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2527031.2527049

[44] Xinning Gui, Yubo Kou, Kathleen H. Pine, and Yunan Chen. 2017. Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4520–4533. https://doi.org/10.1145/3025453.3025891

[45] Sanket Gupta. 2018. Overview of Text Similarity Metrics in Python. Retrieved October 26, 2019 from https://towardsdatascience.com/overview-of-text-similarity-metrics-3397c4601f50

[46] Masahiro Hashizume, Ashraf M Dewan, Toshihiko Sunahara, M Ziaur Rahman, and Taro Yamamoto s. 2012. Hydroclimatological variability and dengue transmission in Dhaka, Bangladesh: a time-series study. *BMC Infectious Diseases* 12, 98 (2012), 9. https://doi.org/10.1186/1471-2334-12-98

[47] Thomas Hofmann. 1999. Probabilistic Latent Semantic Indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Berkeley, California, USA) *(SIGIR '99)*. Association for Computing Machinery, New York, NY, USA, 50–57. https://doi.org/10.1145/312624.312649

[48] Y. Linlin Huang, Kate Starbird, Mania Orand, Stephanie A. Stanek, and Heather T. Pedersen. 2015. Connected Through Crisis: Emotional Proximity and the Spread of Misinformation Online. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 969–980. https://doi.org/10.1145/2675133.2675202

[49] Myung-Hwa Hwang, Shaowen Wang, Guofeng Cao, Anand Padmanabhan, and Zhenhua Zhang. 2013. Spatiotemporal Transformation of Social Media Geostreams: A Case Study of Twitter for Flu Risk Analysis. In *Proceedings of the 4th*

ACM SIGSPATIAL International Workshop on GeoStreaming (Orlando, Florida) *(IWGS '13)*. Association for Computing Machinery, New York, NY, USA, 12–21. https://doi.org/10.1145/2534303.2534310

[50] Mohammad Zahirul Islam, Shannon Rutherford, Dung Phung, Md. Nazim Uzzaman, Scott Baum, M. Mamun Huda, Muhammad Asaduzzaman, Mohammad Radwanur Rahman Talukder, and Cordia Chu. 2018. Correlates of Climate Variability and Dengue Fever in Two Metropolitan Cities in Bangladesh. *Cureus* 10, 10 (2018), e3398. https://doi.org/10.7759/cureus.3398

[51] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2019. Does Transparency in Moderation Really Matter? User Behavior After Content Removal Explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 27. https://doi.org/10.1145/3359252

[52] Ben Jones. 2011. Mixed uptake of social media among public health specialists. Retrieved January 11, 2020 from http://www.who.int/entity/bulletin/volumes/89/11/11-031111.pdf

[53] Nattiya Kanhabua and Wolfgang Nejdl. 2013. Understanding the Diversity of Tweets in the Time of Outbreaks. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13 Companion)*. Association for Computing Machinery, New York, NY, USA, 1335–1342. https://doi.org/10.1145/2487788.2488172

[54] Rathimala Kannan, Kannan Ramakrishnan, and Adedapo Oluwaseyi Ojo. 2019. Social Networking Sites as Communication Tool for Dengue Related Healthcare and Wellness Information. In *Proceedings of the 2nd International Conference on Software Engineering and Information Management* (Bali, Indonesia) *(ICSIM 2019)*. Association for Computing Machinery, New York, NY, USA, 204–207. https://doi.org/10.1145/3305160.3305163

[55] Patty Kostkova, Vino Mano, Heidi J. Larson, and William S. Schulz. 2016. VAC Medi+board: Analysing Vaccine Rumours in News and Social Media. In *Proceedings of the 6th International Conference on Digital Health Conference* (Montréal, Québec, Canada) *(DH '16)*. Association for Computing Machinery, New York, NY, USA, 163–164. https://doi.org/10.1145/2896338.2896370

[56] Moritz U. G. Kraemer, D. Bisanzio, R. C. Reiner, R. Zakar, J. B. Hawkins, C. C. Freifeld, D. L. Smith, S. I. Hay, J. S. Brownstein, and T. Alex Perkins. 2018. Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from Lahore, Pakistan. *EPJ Data Science* 7, 16 (2018), 17. https://doi.org/10.1140/epjds/s13688-018-0144-x

[57] The Lancet. 2019. The Dengue Epidemic in Bangladesh: Risk Factors and Actionable Items. Retrieved March 29, 2020 from https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(19)32524-3.pdf

[58] Chung-Hong Lee, Hsin-Chang Yang, and Shih-Jan Lin. 2015. Incorporating Big Data and Social Sensors in a Novel Early Warning System of Dengue Outbreaks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (Paris, France) *(ASONAM '15)*. Association for Computing Machinery, New York, NY, USA, 1428–1433. https://doi.org/10.1145/2808797.2808883

[59] Kathy Lee, Ankit Agrawal, and Alok Choudhary. 2013. Real-Time Disease Surveillance Using Twitter Data: Demonstration on Flu and Cancer. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Chicago, Illinois, USA) *(KDD '13)*. Association for Computing Machinery, New York, NY, USA, 1474–1477. https://doi.org/10.1145/2487575.2487709

[60] Jiajun Liu, Kun Zhao, Saeed Khan, Mark Cameron, and Raja Jurdak. 2015. Multi-scale population and mobility estimation with geo-tagged Tweets. In *2015 31st IEEE International Conference on Data Engineering Workshops* (Seoul, South Korea). IEEE, Piscataway, New Jersey, USA, 83–86. https://doi.org/10.1109/ICDEW.2015.7129551

[61] May O. Lwin, Santosh Vijaykumar, Owen Noel Newton Fernando, Siew Ann Cheong, Vajira Sampath Rathnayake, Gentatsu Lim, Yin-Leng Theng, Subhasis Chaudhuri, and Schubert Foo. 2014. A 21st century approach to tackling dengue: Crowdsourced surveillance, predictive mapping and tailored communication. *Acta Tropica* 130 (2014), 100 – 107. https://doi.org/10.1016/j.actatropica.2013.09.021

[62] May O. Lwin, Santosh Vijaykumar, Schubert Foo, Owen Noel Newton Fernando, Gentatsu Lim, Chitra Panchapakesan, and Prasad Wimalaratne. 2015. Social media-based civic engagement solutions for dengue prevention in Sri Lanka: results of receptivity assessment. *Health Education Research* 31, 1 (2015), 1–11. https://doi.org/10.1093/her/cyv065

[63] Jim Maddock, Robert Mason, and Kate Starbird. 2015. Using Historical Twitter Data for Research: Ethical Challenges of Tweet Deletions. In *CSCW '15 Workshop on Ethics at the 2015 Conference on Computer Supported Cooperative Work* (Vancouver, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 6.

[64] Jim Maddock, Kate Starbird, Haneen J. Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason. 2015. Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 228–241. https://doi.org/10.1145/2675133.2675280

[65] Ayesha S. Mahmuda, Md. Iqbal Kabir, Sani Tahminad Kenth Engøo-Monsenc, Baizid Khoorshid Riaz, Md. Akram Hossain, Fahmida Khanom, Md. Mujibor Rahman, Md. Khalilur Rahman, Mehruba Sharmin, Dewan Mashrur Hossain, Shakila Yasmin, Md. Mokhtar Ahmed, Mirza Afreen Fatima Lushah, and Caroline O. Buckee. 2019. Megacities as drivers of national outbreaks: the role of holiday travel in the spread of infectious diseases. *null* null, null (2019), null.

https://doi.org/10.1101/737379 arXiv:https://www.biorxiv.org/content/10.1101/737379v1
[66] Tauqeer Hussain Mallhi, Yusra Habib Khan, Azmi Sarriff, and Amer Hayat Khan. 2016. Crimean-Congo haemorrhagic fever virus and Eid-Ul-Adha festival in Pakistan. *The Lancet Infectious Diseases* 16, 12 (2016), 1332–1333. https://doi.org/10.1016/S1473-3099(16)30453-4
[67] Cecilia de Almeida Marques-Toledo, Carolin Marlen Degener, Livia Vinhal, Giovanini Coelho, Wagner Meira, Claudia Torres Codeço, and Mauro Martins Teixeira. 2017. Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level. *PLOS Neglected Tropical Diseases* 11, 7 (2017), 1–20. https://doi.org/10.1371/journal.pntd.0005729
[68] Obaidur Masum. 2019. Women dying more from dengue but disease afflicting more men in Bangladesh. Retrieved January 3, 2020 from https://bdnews24.com/health/2019/08/20/women-dying-more-from-dengue-but-disease-afflicting-more-men-in-bangladesh
[69] Puneet Mathur, Meghna Ayyar, Sahil Chopra, Simra Shahid, Laiba Mehnaz, and Rajiv Shah. 2018. Identification of Emergency Blood Donation Request on Twitter. In *Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task*. Association for Computational Linguistics, Brussels, Belgium, 27–31. https://doi.org/10.18653/v1/W18-5907
[70] John H. McDonald. 2014. *Handbook of Biological Statistic* (3 ed.). Sparky House Publishing, Baltimore, MD.
[71] Sarah F. McGough, John S. Brownstein, Jared B. Hawkins, and Mauricio Santillana. 2017. Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data. *PLOS Neglected Tropical Diseases* 11, 1 (2017), 1–15. https://doi.org/10.1371/journal.pntd.0005295
[72] Raina M. Merchant, David A. Asch, Patrick Crutchley, Lyle H. Ungar, Sharath C. Guntuku, Johannes C. Eichstaedt, Shawndra Hill, Kevin Padrez, Robert J. Smith, and H. Andrew Schwartz. 2019. Evaluating the predictability of medical conditions from social media posts. *PLOS ONE* 14, 6 (06 2019), 1–12. https://doi.org/10.1371/journal.pone.0215476
[73] Paolo Missier, Alexander Romanovsky, Tudor Miu, Atinder Pal, Michael Daniilakis, Alessandro Garcia, Diego Cedrim, and Leonardo Silva Sousa. 2016. Tracking Dengue Epidemics Using Twitter Content Classification and Topic Modelling. In *ICWE 2016 International Workshops on Current Trends in Web Engineering - Volume 9881*. Springer-Verlag, Berlin, Heidelberg, 80–92. https://doi.org/10.1007/978-3-319-46963-8_7
[74] Liesbeth Mollema, Irene Anhai Harmsen, Emma Broekhuizen, Rutger Clijnk, Hester De Melker, Theo Paulussen, Gerjo Kok, Robert Ruiter, and Enny Das. 2015. Disease Detection or Public Opinion Reflection? Content Analysis of Tweets, Other Social Media, and Online Newspapers During the Measles Outbreak in the Netherlands in 2013. *Journal of Medical Internet Research* 17, 5 (2015), e128. https://doi.org/10.2196/jmir.3863
[75] Ivonne Morales, Henrik Salje, Samir Saha, and Emily S. Gurley. 2016. Seasonal Distribution and Climatic Correlates of Dengue Disease in Dhaka, Bangladesh. *The American Journal of Tropical Medicine and Hygiene* 94, 6 (2016), 1359–1361. https://doi.org/10.4269/ajtmh.15-0846
[76] Dhiraj Murthy. 2013. NEW MEDIA AND NATURAL DISASTERS. *Information, Communication & Society* 16, 7 (2013), 1176–1192. https://doi.org/10.1080/1369118X.2011.611815
[77] Pulak Mutsuddy, Sanya Tahmina Jhora, Abul Khair Mohammad Shamsuzzaman, S. M. Golam Kaisar, and Md Nasir Ahmed Khan. 2019. Dengue Situation in Bangladesh: An Epidemiological Shift in terms of Morbidity and Mortality. *Canadian Journal of Infectious Diseases and Medical Microbiology* 2019, 0 (2019), 12. https://doi.org/10.1155/2019/3516284
[78] Meenakshi Nagarajan, Karthik Gomadam, Amit P. Sheth, Ajith Ranabahu, Raghava Mutharaju, and Ashutosh Jadhav. 2009. Spatio-Temporal-Thematic Analysis of Citizen Sensor Data: Challenges and Experiences. In *Web Information Systems Engineering - WISE 2009*. Springer Berlin Heidelberg, Berlin, Heidelberg, 539–553.
[79] Ivo Ngade, Merrill Singer, Olivia Marcus, and José E. Hasemann Lara. 2016. *Washing Away Ebola*. John Wiley & Sons, Ltd, Hoboken, New Jersey, USA, Chapter 8, 157–172. https://doi.org/10.1002/9781118786949.ch8
[80] Elaine O. Nsoesie, Luisa Flor, Jared Hawkins, Adyasha Maharana, Tobi Skotnes, Fatima Marinho, and John S. Brownstein. 2016. Social Media as a Sentinel for Disease Surveillance: What Does Sociodemographic Status Have to Do with It? *PLOS Currents* 7, 8 (2016), 26. https://doi.org/10.1371/currents.outbreaks.cc09a42586e16dc7dd62813b7ee5d6b6
[81] Government of Bangladesh. 2019. Bangladesh Dengue Situation 2019 Health Bulletin - volume: 2, 03 August, 2019. Retrieved November 11, 2019 from https://reliefweb.int/report/bangladesh/bangladesh-dengue-situation-2019-health-bulletin-volume-2-03-august-2019
[82] Government of Bangladesh. 2019. Bangladesh Dengue Situation 2019 Health Bulletin - volume: 3, 10 August, 2019. Retrieved November 11, 2019 from https://reliefweb.int/report/bangladesh/bangladesh-dengue-situation-2019-health-bulletin-volume-3-10-august-2019
[83] Government of Bangladesh. 2019. Dengue Situation, Updated on 22.10.2019. Retrieved October 27, 2019 from https://reliefweb.int/report/bangladesh/dengue-situation-updated-22102019
[84] Ministry of Health and Government of the People's Republic of Bangladesh Family Welfare. 2017. Health Bulletin 2017. Retrieved October 27, 2019 from http://www.dghs.gov.bd/index.php/en/home/4364-health-bulletin-2017

[85] Hyun Jung Oh and Hyegyu Lee. 2019. When Do People Verify and Share Health Rumors on Social Media? The Effects of Message Importance, Health Anxiety, and Health Literacy. *Journal of Health Communication* 24, 11 (2019), 837–847. https://doi.org/10.1080/10810730.2019.1677824
[86] Sunday Oluwafemi Oyeyemi, Elia Gabarron, and Rolf Wynn. 2014. Ebola, Twitter, and misinformation: a dangerous combination? *BMJ* 349 (2014), 2. https://doi.org/10.1136/bmj.g6178
[87] Michael J. Paul and Mark Dredze. 2011. You Are What You Tweet: Analyzing Twitter for Public Health. In *5th International AAAI Conference on Weblogs and Social Media*, Vol. 38. AAAI Publications, Palo Alto, CA, USA, 265–272.
[88] Michael J Paul and Mark Dredze. 2012. A Model for Mining Public Health Topics from Twitter. *Health* 11, 16-16 (2012), 7.
[89] Michael J. Paul and Mark Dredze. 2014. Discovering Health Topics in Social Media Using Topic Models. *PLOS ONE* 9, 8 (2014), 1–11. https://doi.org/10.1371/journal.pone.0103408
[90] Reid Priedhorsky, Dave Osthus, Ashlynn R. Daughton, Kelly R. Moran, Nicholas Generous, Geoffrey Fairchild, Alina Deshpande, and Sara Y. Del Valle. 2017. Measuring Global Disease with Wikipedia: Success, Failure, and a Research Agenda. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1812–1834. https://doi.org/10.1145/2998181.2998183
[91] pypi.org. 2018. tweetscraper 1.2.6. Retrieved September 26, 2019 from https://pypi.org/project/tweetscraper/
[92] Aditya Lia Ramadona, Yesim Tozan, Lutfan Lazuardi, and Joacim Rocklöv. 2019. A combination of incidence data and mobility proxies from social media predicts the intra-urban spread of dengue in Yogyakarta, Indonesia. *PLOS Neglected Tropical Diseases* 13, 4 (04 2019), 1–12. https://doi.org/10.1371/journal.pntd.0007298
[93] Pamela Katz Ressler and Greer Glazer. 2010. Legislative: Nursing's Engagement in Health Policy and Healthcare Through Social Media. *OJIN: The Online Journal of Issues in Nursing* 16, 1 (2010), 491–499. https://doi.org/10.3912/OJIN.Vol16No01LegCol01
[94] Shruti Rijhwani, Royal Sequiera, Monojit Choudhury, Kalika Bali, and Chandra Shekhar Maddila. 2017. Estimating Code-Switching on Twitter with a Novel Generalized Word-Level Language Detection Technique. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, 1971–1982. https://doi.org/10.18653/v1/P17-1180
[95] Carmen Ruiz Vicente, Dario Freni, Claudio Bettini, and Christian S. Jensen. 2011. Location-Related Privacy in Geo-Social Networks. *IEEE Internet Computing* 15, 3 (2011), 20–27. https://doi.org/10.1109/MIC.2011.29
[96] Marcel Salathé and Shashank Khandelwal. 2011. Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. *PLOS Computational Biology* 7, 10 (2011), 1–7. https://doi.org/10.1371/journal.pcbi.1002199
[97] Megha Sharma, Kapil Yadav, Nitika Yadav, and Keith C. Ferdinand. 2017. Zika virus pandemic—analysis of Facebook as a social media health information platform. *American Journal of Infection Control* 45, 3 (2017), 301 – 302. https://doi.org/10.1016/j.ajic.2016.08.022
[98] Sifat Sharmin, Kathryn Glass, Elvina Viennet, and David Harley. 2018. Geostatistical mapping of the seasonal spread of under-reported dengue cases in Bangladesh. *PLOS Neglected Tropical Diseases* 12, 11 (2018), 1–13. https://doi.org/10.1371/journal.pntd.0006947
[99] Luke Sloan and Jeffrey Morgan. 2015. Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter. *PLOS ONE* 10, 11 (2015), 1–15. https://doi.org/10.1371/journal.pone.0142209
[100] Robert Soden. 2018. Crisis Informatics in the Anthropocene: Disasters as Matters of Care and Concern. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3170427.3173027
[101] Robert Soden and Leysia Palen. 2018. Informating Crisis: Expanding Critical Perspectives in Crisis Informatics. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 22. https://doi.org/10.1145/3274431
[102] Roberto C.S.N.P. Souza, Renato M. Assunção, Daniel B. Neill, and Wagner Meira. 2019. Detecting Spatial Clusters of Disease Infection Risk Using Sparsely Sampled Social Media Mobility Patterns. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (Chicago, IL, USA) *(SIGSPATIAL '19)*. Association for Computing Machinery, New York, NY, USA, 359–368. https://doi.org/10.1145/3347146.3359369
[103] Jerry Spiegel, Shannon Bennett, Libby Hattersley, Mary H. Hayden, Pattamaporn Kittayapong, Sustriayu Nalim, Daniel Nan Chee Wang, Emily Zielinski-Gutiérrez, and Duane Gubler. 2015. Barriers and Bridges to Prevention and Control of Dengue: The Need for a Social–Ecological Approach. *EcoHealth* 2, null (2015), 273–290. https://doi.org/10.1007/s10393-005-8388-x
[104] S. Sridhar, P. Gautret, and P. Brouqui. 2015. A comprehensive review of the Kumbh Mela: identifying risks for spread of infectious diseases. *Clinical Microbiology and Infection* 21, 2 (2015), 128 – 133. https://doi.org/10.1016/j.cmi.2014.11.021

[105] StatCounter Global Stats. 2019. Social Media Stats Bangladesh. Retrieved October 3, 2019 from https://gs.statcounter.com/social-media-stats/all/bangladesh/2019

[106] Ravi Suggala. 2019. A survey on prediction approaches for epidemic disease outbreaks based on social media data. *International Journal of Advanced Trends in Computer Science and Engineering* 8, null (2019), 897–908. https://doi.org/10.30534/ijatcse/2019/86832019

[107] Hannes Taubenböck, Jeroen Staab, Xiao Zhu, Christian Geiß, Stefan Dech, and Michael Wurm. 2018. Are the Poor Digitally Left Behind? Indications of Urban Divides Based on Remote Sensing and Twitter Data. *ISPRS International Journal of Geo-Information* 7, 8 (2018), 304. https://doi.org/10.3390/ijgi7080304

[108] D. Teo, L. C. Ng, and S. Lam. 2009. Is dengue a threat to the blood supply? *Transfusion Medicine* 19, 2 (2009), 66–77. https://doi.org/10.1111/j.1365-3148.2009.00916.x

[109] C. Lee Ventola. 2014. Social Media and Health Care Professionals: Benefits, Risks, and Best Practices. *Pharmacy and Therapeutics* 39, 7 (2014), 491–499.

[110] Venturini and Richard Rogers. 2019. "API-Based Research" or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach. *Digital Journalism* 7, 4 (2019), 532–540. https://doi.org/10.1080/21670811.2019.1591927

[111] Helena Webb, Marina Jirotka, Bernd Carsten Stahl, William Housley, Adam Edwards, Matthew Williams, Rob Procter, Omer Rana, and Pete Burnap. 2017. The Ethical Challenges of Publishing Twitter Data for Research Dissemination. In *Proceedings of the 2017 ACM on Web Science Conference* (Troy, New York, USA) *(WebSci '17)*. Association for Computing Machinery, New York, NY, USA, 339–348. https://doi.org/10.1145/3091478.3091489

[112] WHO. 2006. Flooding and communicable diseases fact sheet. Retrieved January 3, 2020 from https://www.who.int/hac/techguidance/ems/flood_cds/en/

[113] WHO. 2011. Vector surveillance. Retrieved November 15, 2019 from https://www.who.int/denguecontrol/monitoring/vector_surveillance/en/

[114] WHO. 2019. Health information and knowledge management. Retrieved November 15, 2019 from http://www.searo.who.int/bangladesh/areas/healthinformation/en/

[115] WHO. 2020. Dengue and severe dengue. Retrieved March 30, 2020 from https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

[116] Wikipedia. 2016. List of hospitals in Bangladesh. Retrieved January 4, 2020 from https://en.wikipedia.org/wiki/List_of_hospitals_in_Bangladesh

[117] Volker Wulf, Kaoru Misaki, Meryem Atam, David Randall, and Markus Rohde. 2013. "On the Ground" in Sidi Bouzid: Investigating Social Media Use during the Tunisian Revolution. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1409–1418. https://doi.org/10.1145/2441776.2441935

[118] Munesada Yamakawa, Miyuki Yamauchi, Minoru Nidaira, Tomoyuki Azuma, Tadashi Nakasone, Fumie Ando, Takahiro Hayamizu, Saki Kinjo, and Yuuki Nakamura. 2017. Enhanced Public Health Surveillance for the Sixth Worldwide Uchinanchu Festival Conducted by the Okinawa Prefectural Government, Japan. *Scientific Research* 5, 9 (2017), 106–115. https://doi.org/10.4236/jbm.2017.59010

[119] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. A Biterm Topic Model for Short Texts. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery, New York, NY, USA, 1445–1456. https://doi.org/10.1145/2488388.2488514

[120] Sanni Yaya, Olalekan A. Uthman, Agbessi Amouzou, and Ghose Bishwajit. 2018. Mass media exposure and its impact on malaria prevention behaviour among adult women in sub-Saharan Africa: results from malaria indicator surveys. *Global Health Research and Policy* 3, 20 (2018), 9. https://doi.org/10.1186/s41256-018-0075-x

[121] Douglas Yeung. 2018. Social Media as a Catalyst for Policy Action and Social Change for Health and Well-Being: Viewpoint. *Journal of Medical Internet Research* 20, 3 (2018), e94. https://doi.org/10.2196/jmir.8508

[122] Elad Yom-Tov, Diana Borsa, Ingemar J Cox, and Rachel A McKendry. 2014. Detecting Disease Outbreaks in Mass Gatherings Using Internet Data. *Journal of Medical Internet Research* 16, 6 (2014), e154. https://doi.org/10.2196/jmir.3156

[123] Qingyu Yuan, Elaine O. Nsoesie, Benfu Lv, Geng Peng, Rumi Chunara, and John S. Brownstein. 2013. Monitoring Influenza Epidemics in China with Search Query from Baidu. *PLOS ONE* 8, 5 (2013), 1–7. https://doi.org/10.1371/journal.pone.0064323

[124] Michael Zimmer. 2010. Is it ethical to harvest public Twitter accounts without consent? Retrieved March 21, 2020 from https://www.michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/

Received January 2020; accepted March 2020