CovidSens: A Vision on Reliable Social Sensing based Risk Alerting Systems for COVID-19 Spread

Md Tahmid Rashid, Dong Wang
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
{mrashid, dwang5}@nd.edu

Abstract—With the spiraling pandemic of the Coronavirus Disease 2019 (COVID-19), it has becoming inherently important to disseminate accurate and timely information about the disease. Due to the ubiquity of Internet connectivity and smart devices, social sensing is emerging as a dynamic sensing paradigm to collect real-time observations from online users. In this vision paper we propose CovidSens, the concept of social-sensing-based risk alerting systems to notify the general public about the COVID-19 spread. The CovidSens concept is motivated by two recent observations: 1) people have been actively sharing their state of health and experience of the COVID-19 via online social media, and 2) official warning channels and news agencies are relatively slower than people reporting their observations and experiences about COVID-19 on social media. We anticipate an unprecedented opportunity to leverage the posts generated by the social media users to build a real-time analytic system for gathering and circulating vital information of the COVID-19 propagation. Specifically, the vision of CovidSens attempts to answer the questions of: how to track the spread of the COVID-19? How to distill reliable information about the disease with the coexistence of prevailing rumors and misinformation in the social media? How to inform the general public about the latest state of the spread timely and effectively and alert them to remain prepared? In this vision paper, we discuss the roles of CovidSens and identify the potential challenges in implementing reliable social-sensing-based risk alerting systems. We envision that approaches originating from multiple disciplines (e.g. estimation theory, machine learning, constrained optimization) can be effective in addressing the challenges. Finally, we outline a few research directions for future work in CovidSens.

Index Terms—Social sensing, COVID-19, coronavirus, disease tracking, real-time, information distillation.

I. INTRODUCTION

Due to the pervasion of Internet connectivity and smart devices, social sensing is an gradually escalating to a new sensing paradigm that utilizes observations by humans and devices on their behalf to obtain information about the physical world [1]. In this vision paper we present CovidSens, the notion of a real-time risk alerting system based on social sensing to guide situational awareness and intervention motives for Coronavirus Disease 2019 (COVID-19) spread. According to the most recent statistics, there are more than 435,000 confirmed cases of COVID-19 and above of 14,800 death spread across 50 states in US [2], [3]. Most of the above cases happened within one week's time (i.e., between March 26, 2020 and April 01, 2020) and the current trend seems to be ever increasing [3]. As the outbreak of COVID-19 progresses, circulating information about the spread in an accurate and timely manner has grown ever important. However, with heightening uncertainty and commotion among the general public, the communication of timely and accurate information to intended recipients is a challenging task. While official warning channels and news agencies have served an active role in informing the public about the spread, they often fall short in terms of pace. It is apparent that the official warning channels and news media take a while to confirm and disseminate the information regarding the outbreak of a new disease [4]. By contrast, information propagation across the social media and crowd-sensing platforms is inherently faster than traditional news media [4]. For example, during the 2013 Boston Marathon Bombing, news about the first bomb explosion and the arrest of the suspect were posted in Twitter several minutes before news agencies made announcements [5], [6]. After the onset of Cholera outbreak in Haiti in 2010, the knowledge regarding the outbreak was first obtained through social media, which occurred weeks before officials confirmed the case of the outbreak [7]. Such cases exemplify the importance of social media during emergency scenarios such as now during the COVID-19 outbreak.

The CovidSens concept is thus motivated by two observations during this global crisis of COVID-19. Firstly, people tend to actively convey their state of health and experience of the virus via online social media since the onset of the COVID-19. For instance, at one given day, 6.7M people talked about coronavirus on social media [8]. Secondly, people report their observations on social media relatively faster than official warning channels and news agencies make formal announcements. As such, knowledge contribution and discovery through social sensing may offer more effective news transmission [4]. Given this premise, we perceive an unprecedented opportunity to leverage the posts generated by the social media users to build a complete analytics framework for gathering and circulating vital information of the COVID-19 propagation.

Let us consider a few tweets posted during the course of the COVID-19 spread across the US in Figure 1. These tweets express the experiences and observations of individuals about the COVID-19. If such tweets could be utilized to identify regions affected by COVID-19 and the rate of spread of the virus, it might potentially expedite the alleviation of the adverse effects of the virus. In addition, by analyzing the location

and movement data from smartphones and social media posts to detect crowds or mass gatherings while respecting user privacy, government agencies and the mass public could be informed about the more risk-prone areas of a city during the COVID-19 outbreak [9]. This could potentially help to divert people away from more crowded locations and hence reduce the spread of the disease.

Figure 1. Tweets posted during the COVID-19 outbreak

While the CovidSens vision promises opportunities for a robust information distillation as well as risk alert service for the COVID-19 spread, several technical challenges exist in the way of building such a system that would spontaneously gather and distribute real-time development of the disease to the general public. In contrast to traditional disaster response systems (e.g., for floods or forest fires), one unique goal of CovidSens is to obtain knowledge of the dynamics of the disease spread (e.g., inferring the stages of the disease among people). The first challenge is, therefore, to build a social sensing data collection platform that is able to spontaneously obtain the relevant social signals about symptoms, cases, and fatalities of COVID-19 from the online social media users. The second challenge lies in developing reliable data analysis models that can extract the credible information of the disease spread from the noisy, sparse, and unstructured social data contributed by unvetted human sources such as the three tweets presented in Figure 1. The third challenge exists in handing the huge volumes of social data about the COVID-19 outbreak that varies widely (e.g., across text, image, video, and audio data). The fourth challenge is to circulate the extracted information about the disease spread to the general public in a timely and efficient manner so that they can plan their actions accordingly. The fifth challenge lies in designing an effective alert system that considers the human aspect of the problem (i.e., handling people's reaction to alerts like fear, concern, or ignorance). The sixth challenge is combating the misinformation spread in the social media where people tend to report rumors or falsified facts of the COVID-19 spread.

The CovidSens aims to overcome the above limitations by providing a more reliable and timely COVID-19 monitoring and alerting system for the mass population based on social sensing. We envision an information retrieval and dispatching system for the general public based on data derived from multiple sources (e.g., social media, crowdsourced platforms, Unmanned Aerial Vehicle (UAV)) to quickly and effectively monitor the spread of the COVID-19 using combinations of smartphone applications, UAVs, message boards, or other modes of information dispersal. We expect this service to be important and useful for people who live in or travel to the affected areas, allowing them to take special precautions and be well prepared. The successful development of such systems can potentially help both authorities and general public respond more quickly and efficiently to COVID-19 and eventually help save more lives.

We acknowledge the potential to employ interdisciplinary techniques from estimation theory, online social media analysis, machine learning, AI and mobile phone applications to develop effective CovidSens systems. Research along the realm of CovidSens is important because the COVID-19 is spreading rapidly in many countries worldwide and a timely alerting system that explores the rich real-time information streaming on social media is yet to be developed. The results of this research can pave the way for studying and tackling COVID-19 around the world.

The rest of the paper is organized as follows. In Section II, we discuss a few state-of-the-art works in the direction of CovidSens. In Section III, we explore potential real-world applications of CovidSens. We identify the a few likely challenges in implementing a successful CovidSens system in Section IV. Afterwards in Section V, we highlight a set of research directions for future work aligning with CovidSens to contain the COVID-19 spread. Finally, we conclude our vision of CovidSens in Section VI.

II. RELATED WORK

A. Social Sensing

Social sensing is rapidly progressing as a pervasive sensing paradigm where humans are used as sensors to attain situational awareness about the physical world [4]. Examples of social sensing applications include predicting poverty in developing countries [10], studying human mobility in urban areas [11], identifying traffic abnormalities [12], tracking social unrest [13] and disasters [14], classifying the urban land usage [15], detecting wildfire [16], identifying the point of interests in cities [17], and performing post-disaster damage assessments [18]. A comprehensive survey of social sensing schemes is provided in [19]. Zhang *et al.* developed a scalable approach to obtain data veracity in social sensing [20]. Xu *et al.* developed a framework for semantic and spatial analysis of urban emergency events using social media data [21]. Zhang *et al.* presented a constraint-aware truth discovery model to detect dynamically evolving truth in social sensing [22]. More recently, there is an advent of social-media-driven drone sensing (SDS) approaches that address the data reliability issue of social sensing by integrating social signals with physical UAVs [23]–[25]. While existing social sensing approaches aim to provide pervasive sensing, they are not tailored specifically to monitor the COVID-19 outbreak. Compared to traditional social sensing applications, CovidSens not only requires an

inference of the data veracity, but also how the COVID-19 outbreak can progress across regions based on indications from social media posts (e.g., posts about crowded subways could indicate high risk of COVID-19 risk spread). Thus, it remains a critical task to develop a reliable social sensing model that can accurately monitor the COVID-19 spread.

B. Disease Outbreak Investigation

In recent times, disease tracking based on epidemiological data has been an important avenue of research. Several studies have independently explored the feasibility of using social media and crowdsensing for detection, tracking, and analytics of contagious disease outbreaks [26], [27]. For example, Google launched a real-time influenza surveillance system, namely Google Flu Trends [28], to monitor influenza spread by analyzing search terms related to illness symptoms. Kalogiros *et al.* developed Allergymap, a crowdsensing-based disease identification system for allergen season onsets and allergy patient stratification [29]. Krieck *et al.* studied the possibility of analyzing Twitter data for infectious disease surveillance [30]. Chester *et al.* [31] carried out bacterial outbreak investigation based on web forum posts about sick participants from a bike race. Despite the advances in disease monitoring techniques, current schemes have not been designed to handle the exponential progression of the COVID-19 pandemic and provide reliable risk alert in the context of CovidSens. Therefore, it entails a more rapid information distillation and processing system that can track the COVID-19 spread in real-time.

C. Automated Disease Warning and Alert Systems

While traditional health systems play an important role in alerting the general public about infectious diseases, their slow information progression have necessitated the adoption of automated warning and alert systems [26]. Brownstein *et al.* contributed a few early works in this domain by developing: i) a series of interactive websites, HealthMap and Flu Near You [26], [32], and ii) a smartphone application called Outbreaks Near Me [33] to present vital information about outbreaks of various illnesses around the world. Toda *et al.* explored the effectiveness of a text-messaging system for notification of disease outbreaks in Kenya [34]. Yu *et al.* developed ProMED-mail, an early warning system for emerging diseases [35]. Carter studied the possibility of a tweet-based information dispersal system to facilitate the containment of Ebola in Nigeria [36]. The above approaches are known to provide disease warning with reasonable effectiveness. However, it is an even more challenging task to develop a real-time COVID-19 spread indicator for CovidSens that uses both social-media and crowdsourced data, and also transmit the news of the spread to the general public in real-time.

D. COVID-19 Spread Monitoring

With the emergence of the COVID-19 outbreak, several streams of research have introduced methods to monitor the COVID-19 propagation. Sun *et al.* [37] proposed the first study that harnesses crowdsourced data from several social media sources to monitor the COVID-19 spread. Schiffmann [38] developed an informative web portal that aggregates news from myriads of news sources to present latest information on COVID-19 spread. The Johns Hopkins Center for Systems Science and Engineering (JHU CSSE) developed an interactive online dashboard to track and present worldwide reported cases of COVID-19 in real-time [39]. An online community of international students and professionals, called 1point3acres, developed a web-based real-time COVID-19 news aggregator to track the state of the spread in the US and Canada [40]. A mobile app has been developed by the Singaporean government to leverage crowdsourced information to locate community transmission of COVID-19 [41]. A key drawback of the above tools is that they possess partial autonomy, requiring some degree of manual efforts to validate the information of the COVID-19 spread before presenting them online [38], [40]. During this evolving COVID-19 outbreak, delays are undesirable. Therefore, a significant limitation exists in existing approaches to spontaneously track the COVID-19 propagation and disseminate the information to the end users.

III. REAL-WORLD APPLICATIONS

In this section, we highlight a few probable applications in real-world scenarios aligning with the CovidSens vision.

A. Social-media-driven Disease Spread Indicator

In a social-media-driven disease spread indicator (SDSI), social media posts related to COVID-19 are analyzed to attain the state of the spread [37]. An example of an SDSI architecture is illustrated in Figure 2. Initially, a real-time Twitter data crawler engine collects tweets indicating public opinions about the disease. The tweets are subsequently filtered and labelled into discrete categories based on the topics of discussions. A few examples of these topics can be: i) what regions are being frequently reported to be infected; ii) the time between people first talking about COVID-19 symptoms to deciding to be tested (i.e., how long the virus takes to show effect in people) [37], iii) which age of people are expressing about symptoms the most; iv) how rapidly authorities are responding to the stimuli; and v) whether people are talking about other people they know getting recovered [37], [42]. Afterwards the labeled Twitter data are passed to a tweet analytics and training engine on a backend server. Specifically, the backend server will construct a clean and timely events summary about the COVID-19 spread by distilling relevant and reliable information from the massive amount of noisy, unstructured, and unvetted data feeds. Lastly, a website or smartphone app will interact with end users to provide them warnings or alerts about the disease spread in their vicinity based on their queries. The analytics engine jointly analyzes the data veracity, source reliability, observation bias (e.g., under vs over estimation), as well as the likelihood of large-scale havoc launched by malicious users on social media using novel estimation theoretic, machine learning and AI techniques.
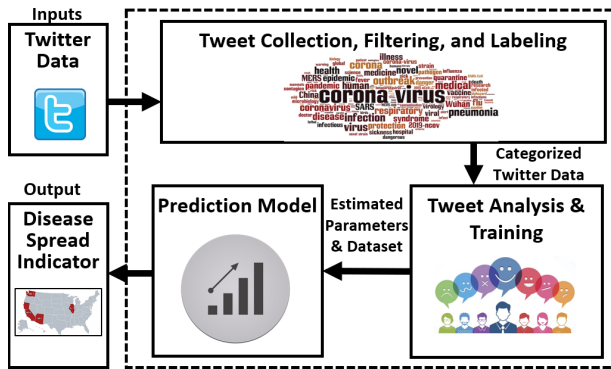
Figure 2. Overview of an SDSI system

Figure 3. Overview of a CDT system

B. Crowdsensing-based Disease Tracking

Crowdsensing-based disease tracking (CDT) involves sensor networks and groups of people, with mobile devices capable of sensing, collectively sharing disease related information (e.g., early symptoms, nearby infected persons, deciding to self-quarantine) [37], [43]. CDT is fueled by the observation that individuals tend to proactively volunteer in contributing data about the COVID-19 spread using their smartphones, wearables, or other devices with sensors and connectivity [37]. In contrast to SDSI, CDT is relatively less pervasive and requires active participation of people and physical sensors. However, in return the data is less noisy and is hence more reliable. Figure 3 shows an example of a representative CDT system. A CDT may typically incorporate three main components. The first component is a data collection platform consisting of a network of users with a custom smartphone application to log data and a set of internet-of-things (IoT) devices (e.g., smart heart-rate monitors, activity trackers, thermal scanners). The smartphone application interacts with users and allows them to actively contribute their reports on the COVID-19 if they are willing to. If the users choose to input data, the app lets the users configure at what granularity (e.g., state, county, street, or N/A) they feel comfortable to share their location information. The second component is an analytics framework that applies relevant statistical analysis and machine learning techniques on the obtained data to infer probable regions of infection and safe zones [33], [43]. The third component is a smartphone application on the end users' mobile phones to visually represent the analyzed geospatial distribution of the inferred regions [33]. The app can obtain the needed information from the backend server based on the users' queries (e.g., checking the risk level of a particular area of interest). In most cases the data collection and representation is carried out in the same smartphone application [33]. Sun *et al.* proposed one of the earliest crowdsourcing based COVID-19 outbreak detection system [37]. The Singaporean and South Korean governments have launched mobile apps that utilize crowdsourced data to trace community transmission of the COVID-19 [41].
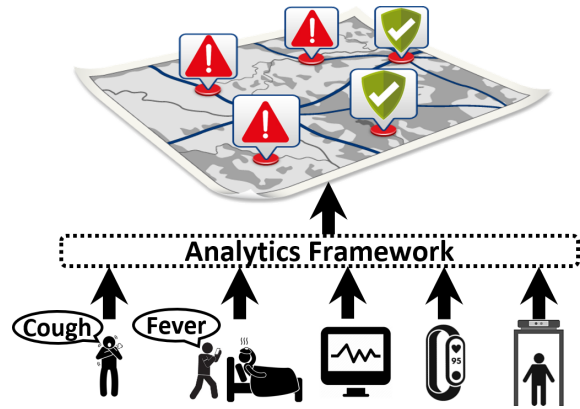
C. UAV-based Health Surveillance and Alerting

The urgency of the COVID-19 outbreak has necessitated new dimensions for UAV-based health surveillance and alerting (UHSA) systems [44]. With the help of onboard sensors (e.g., cameras, microphones), UAVs are able to gather intelligence remotely during a disease pandemic scenario where human patrol teams and ground units cannot operate due to risks of getting infected. For instance, UAVs can assist in detecting unwanted crowds of people along locked down areas of a city [44]. Figure 4 demonstrates a representative UHSA model for mitigating the COVID-19 spread. The UHSA system responds to emergency requests by individuals through social media posts about unnecessary mass gatherings. Afterwards, the data is gathered in a backend server and processed using social sensing approaches based on statistics and machine learning for analyzing the truthfulness of the data. The information is then updated across nearby regions by raising verbal alerts through speakers installed on the UAVs. UAVs are also dispatched out to different areas of a city to spontaneously scan and obtain situational awareness about the region. Using the onboard sensors UHSA detects if people are breaking the rules during the lock down situation (e.g., by roaming outside, gathering in crowds). The framework may also locate the availability of critical supplies using the UAVs' cameras (e.g., open pharmacy, grocery stores) based on the social media posts. Using the onboard speakers, the people breaking the rules are alerted to return home. One real-world example of UHSA during the COVID-19 ordeal is in California, USA where the law enforcement officials have resorted to utilizing drones for patrolling the state of California during the ongoing lock-down situation [45]. During the COVID-19 crisis in China, UAVs have served multiple roles including post-epidemic aerial evaluation, alerting, and relief distribution to affected regions [46].

IV. RESEARCH CHALLENGES AND OPPORTUNITIES

In this section, we present a set of prevalent research challenges and opportunities in the development of an effective CovidSens framework.
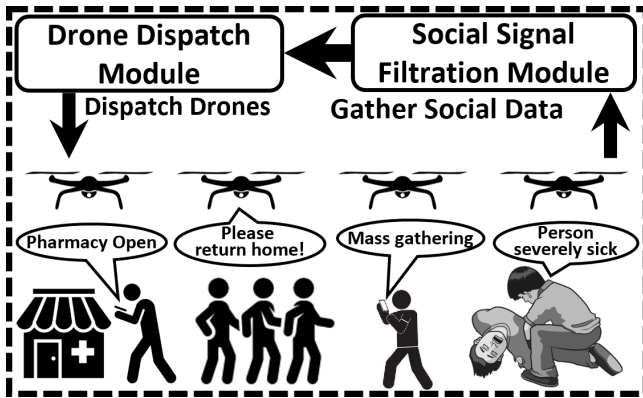
Figure 4. Overview of an UHSA system

A. Data Collection Challenge

During the onset of rampant disease outbreaks like COVID-19, the primary objective of a CovidSens system is to collect information from the general public. However, several difficulties prevail to locate and obtain the relevant posts related to the COVID-19 spread. For instance, while conducting simple keyword based searches on obtained social media data, the desired keywords may indicate various other unwanted things (e.g., while the term "sick" is generally used to indicate people who are not doing well, it may also be used to express sarcasm by certain people). Several recent studies focused on mitigating this issue of data discovery by replacing simple keyword based searches with singular value decomposition (SVD) driven K-means clustering [47] and recurrent neural network (RNN) based textual labeling process [48]. However, such methods still lag behind human perception in terms of accurately scanning for relevant input data. Thus, obtaining a collection of relevant social media data that directs to the right set of information remains an arduous task. Moreover, a great portion of social media data may eventually turn out to be redundant (e.g., retweets) or simply rephrased from a single original post [49]. On top of that, a good amount of social media data is observed to be transient and perishable. For example, people may delete their previous posts and online repositories (i.e., Twitter and Facebook servers) hosting the posts may take them down for undisclosed reasons. In addition to that, social media APIs such as Twitter often impose various rate limitations which can heavily impede the data collection during disease outbreaks [50]. The data collection process for COVID-19 therefore necessitates a tool that can locate, obtain, and store the relevant information from users in real-time across social media channels.

B. Data Reliability Challenge

The concept of CovidSens is centered around the noisy and unreliable data generated by the unknown human sources on the social media [19], [51]. One important task while harnessing social media for CovidSens is to extract trustworthy information from unreliable human sources with unknown source reliability [1], [25]. We define this as the *data reliability*

challenge in social sensing. Several truth discovery solutions have been developed to mitigate the data reliability problem. For instance, Wang et al. presented a framework to jointly estimate the reliability of data sources and the correctness of the reported measurements in social media posts using approaches from estimation theory [1], [52]. Zhang et al. built upon the previous framework to address the scalability and physical constraint challenges and employed the improved schemes to real-world social sensing applications [20], [22], [53]. Yin et al. developed Truth Finder, a probabilistic algorithm using iterative weight updates to improve the quality of the data in social sensing [54]. While great efforts have been made on developing reliable social sensing solutions, certain limitations hinder these solutions from being applied in CovidSens to track COVID-19. One drawback of traditional social sensing schemes is that they solely rely on the noisy social media data and there no external means of validating the credibility of the input data during the COVID-19 epidemic [22]. Existing methods are also not tailored towards disease outbreak detection, which may lead to prediction of false cases of COVID-19. For example, a person simply posting a symptom of breathing difficulty may not necessarily suffer from COVID-19. It may be required to analyze other traits of the patient based on earlier posts. Hence, it remains an unresolved challenge in CovidSens to develop reliable social sensing models that can explore the uncertainty in the input data and extract reliable signals.

C. Data Modality Challenge

While data collection is an intrinsic challenge in using social sensing for tracking the COVID-19 spread, a greater difficulty exists in processing the rapidly generated incoming signals consisting of multitudes of features or dimensions [55]. This challenge is identified as *data modality* in social sensing where large amounts of unfiltered and unstructured data with multiple modalities need to be processed [56]. Specifically, data modality refers to the different variety or types of data prevalent in the social media such as text, image, location, audio, and video [57]. Moreover, each type can further encompass different dimensionality as well which makes the data modality challenge even harder. Examples of dimensionality in CovidSens can range along reports of: i) proximity to infected locations, ii) number of suspected cases, iii) number and types of symptoms, iv) intensity of symptoms (i.e., mild, moderate, or severe), v) recovery rate, vi) death rate, and vii) number of self-quarantined cases. Recent social sensing tools primarily focus on analyzing the text data in social media [58]. This trend is advocated by the fact that image data processing involves heavy computation requirement [59]. Consequently, existing methods do not focus on fusing multiple types of data which may potentially generate richer detection of COVID-19 propagation. For example, a person may tweet about having COVID-19, but based on an image posted with the tweet it may turn out that the person's symptoms have actually resulted from an allergic reaction instead [60]. Fusing text with other data such as image and location data may po-

tentially yield more accurate prediction of the COVID-19 spread. Therefore, given the sheer volumes of multi-modal data generated by the social media users about the COVID-19 outbreak, solutions need to be developed to efficiently utilize the different modality of data. Moreover, since multi-modal data processing intrinsically demands greater compute power, care must be given to strike efficient an trade-off between detection accuracy and computational complexity. A set of unsolved questions springing from the data modality challenge in CovidSens are: i) how to efficiently fuse the different types of social media data related to COVID-19 in to one unified data stream? ii) How to design algorithms to process a wide variety of social data in real-time for an accurate prediction of the COVID-19 spread? iii) How to speed up the analysis of multi-modal data for faster COVID-19 spread detection by distributing the computation across multiple devices?

D. Location Data Scarcity Challenge

One recurring issue in social sensing is the user privacy whereby the personal information of the online users remains at risk of falling into the wrong hands [61]. Geo-location data shared by users can also be used to expose other private information as well (e.g., ethnicity, race, financial status) which social media users do not typically consent to share and are also not required by CovidSens applications. Thus, it has been observed that due to the concern of one's location and private information being exposed, many social media users tend to not share their location information while reporting their observations in the social media [12]. For example, in an independent study involving data collection for disaster related tweets, it was found that less than 10% of the tweets were actually geo-tagged (i.e., contained geographical location of the users). As such, CovidSens applications that heavily rely on the location metadata from the social media posts to provide inference of the COVID-19 spread may under-perform when the number of geo-tagged social media are scarce. Recent literature has explored methods to work around this issue by exploiting spatiotemporal social constraints for location inference from social media posts [62]. However, such uni-dimensional approaches that rely on solely on the content of the social media posts may result in high estimation errors for the inferred locations. In order to precisely track the progress of the COVID-19 propagation, it is imperative to obtain the exact locations of the surges. Consequently, it is a challenge in CovidSens applications to design a solution that can mitigate the data scarcity issue which may eventually yield better sensing results for tracking the COVID-19 spread.

E. Timely Presentation Challenge

With the rapidly evolving circumstances during the COVID-19 outbreak, it is critical to present the information of the disease spread to the end users in a timely manner. This necessitates an information presentation system that can both process as well as present data of the disease propagation in real-time and keep people alerted. In the recent past, several methods have been implemented to present disease outbreak updates to the mass through means of interactive websites [26], [32]. However, such methods of information distribution and collection solely rely on aggregating knowledge from different news portals and information websites which can lead to potential delays in alerting people about the most recent situation [4]. Due to their structured nature of information crawling and collating, existing web-based techniques cannot be directly applied to social sensing which encompasses unstructured and noisy social data [55]. In addition to that, websites and smartphone applications rely on the constant availability of both the Internet and a smart device, either of which may not be available in all circumstances. Thus, vital information may not reach to all sectors of the population, especially with the elderly and less tech savvy individuals without access to computers and smart devices. Based on these grounds, it remains an open question in CovidSens on how to develop a reliable yet efficient mechanism that can rapidly deliver important messages and information regarding the COVID-19 spread to all segments of the population.

F. Human Factor Challenge

One important aspect to consider while dealing with social signals in CovidSens is the human component. Given the intensifying concerns and panic among the general public during the COVID-19, we acknowledge that people can be overly emotional, sensational, or biased in expressing their opinions in the social media or the crowdsensing applications [63]. Such behaviour can potentially trigger misrepresented or misinterpreted observations and thus yield erroneous disease tracking results. Based on the above concerns, one critical challenge stemming from the human aspect of social sensing can be on deciding how to handle the mood of the population while containing the public concern at desirable levels. Moreover, it is imperative to study the human component closely and model how people react to the information presented to them through the warning and alert systems in CovidSens. Some individuals may turn out to be excessively sensitive and thus care must be taken so as not to develop the grounds for unnecessary panic or civil unrest. For example, during the Ebola epidemic in Liberia in 2014, riots broke out among the residents when officials raised alarms of the outbreak [64]. On the other extreme of the spectrum, we also acknowledge that a certain proportion of the population have a tendency to be oblivious of the circumstances, neglect warnings, and remain excessively calm during this outbreak situation. The challenge of CovidSens is to strike a smooth balance between raising attention and providing assurance: at one end we need to calm people down while informing them of the situation but at the same time we also need to send out the message to remain well-prepared.

G. Misinformation Spread Challenge

With the heightening concern of the COVID-19 spread, just as social media has served as a platform for attaining information, it has also served as the venue for sprouting misinformation. Due to the increased adoption of social

sensing as a news source, misinformation spread on social media has remained an inevitable issue [54]. This has caused social media giants such as Facebook and Google to conduct worldwide campaigns to fight the propagation of fake news [65]. Figure 5 illustrates a collection of tweets referring to misinformation during the COVID-19 outbreak. The World Health Organization (WHO) has been forced to reallocate considerable resources to combat swathes of misinformation like these, which may potentially hinder COVID-19 monitoring efforts [66]. This phenomenon has been classified by WHO as an 'infodemic' [66]. Social sensing tools, otherwise known as *truth discovery* algorithms, are known to under-perform in the presence of widespread misinformation, which is common during disease outbreak scenarios. One obvious measure to address this issue is to acquire ground truth for validating the source reliability and event correctness. However, obtaining such ground truth is delay prone since it requires a significant amount of manual effort, but most importantly it is impractical during the course of virus breakouts where people should restrict locomotion and contact with other people. Therefore, it remains a critical challenge in CovidSens to construct an effective mechanism that can identify and isolate the misinformation spread to generate trustworthy social signals indicating the COVID-19 spread.

Figure 5. Tweets indicating fake news

V. ROAD-MAP FOR FUTURE WORK

In this section, we discuss a few potential directions for future work in the realm of CovidSens.

A. Uncertainty Quantification in CovidSens

We note that CovidSens relies on noisy and uncertain social-sensing data generated by unvetted data sources to monitor the COVID-19 spread. Thus, one domain for future work can be to mitigate the data reliability challenge for CovidSens applications. We observe that existing social-sensing tools or *truth discovery* algorithms mainly prioritize on the data veracity or source reliability from the social media data. However, in a social-media-driven COVID-19 spread indicator application, the estimation confidence of a reported event's veracity is also crucial [67]. Consequently, it is important to determine the confidence level with which the COVID-19 propagation is predicted. For example, an inferred age demography with a low estimation confidence can easily lead to an erroneous conclusion on which age of people are most likely to be affected by COVID-19. In particular, further research can focus on rigorously quantifying the uncertainty of output results to evaluate and enhance the performance of the truth discovery algorithms. While the uncertainty quantification is well-studied in statistics and estimation theory, it is mostly overlooked in existing social sensing solutions since the performance of truth discovery algorithms are hard to inspect and humans are more likely to generate the claims with different degrees of uncertainty (e.g., affirmative assertions versus pure guesses) [68], [69]. Based on this, one probable research direction is to develop a method to determine the confidence levels of detection by quantifying the uncertainty of the results in CovidSens applications.

Current literature on statistical analysis discusses principled approaches based on estimation theory. A set examples of techniques to quantify the uncertainty of the estimation results of the truth discovery algorithms are maximum likelihood estimation (MLE) and Cramer-Rao lower bounds (CRLB) [53], [67]. While these methods have been tested to operate optimally to provide the desired uncertainty quantification, it stills remain a critical challenge to formulate the truth discovery problems in CovidSens in a mathematically tractable way that would allow the uncertainty estimation tools to be applied upon. We envision that theories from multiple disciplines would be leveraged to cater to the uncertainty quantification problem in the CovidSens applications.

B. Rumor Suppression and Fake News Detection

One direction for future work for CovidSens is to combat the misinformation propagation challenge. Therefore, rumor suppression and fake news detection are indispensable for COVID-19 related misinformation spread containment. We acknowledge that rumors and misinformation in social media originates from the behaviour of individuals sharing what others share [70]. Thus, it is beyond the scope of machine intelligence alone to contain the spread of rumors and misinformation entirely. Based on these premises, a few potential research questions can be: i) how to develop techniques that incorporate human intelligence along with machine intelligence to more accurately identify the rumors from true information about the COVID-19 spread? ii) How to investigate and identify the origin behind misinformation sharing from the social media posts? iii) How different demography (e.g., age groups, gender classes) react to misinformation about COVID-19 spread and how to utilize this knowledge to combat the misinformation propagation?

Several existing literature has proposed different fact determining techniques for analyzing and detecting falsified claims and rumors on social media using: i) Bayesian-based heuristic algorithms [54], ii) analyzing textual evidence with associated images [71], and iii) considering physical constraints and temporal dependencies of the evolving truth [72]. One new domain of research focuses on unifying the collective strengths of human intelligence (HI) and artificial intelligence (AI) to screen out misinformation in the social media [73]. Such approaches utilize HI-based crowdsourcing platforms such

as Amazon Mechanical Turk (MTurk) in combination with existing deep neural networks (DNNs) and machine learning techniques, and can be used to classify social media posts about COVID-19 as veracious or falsified [73].

C. Mesh Network for News Aggregation and Circulation

A stream of potential research can focus around mitigating the data collection and timely presentation challenges in CovidSens applications. In order to obtain information, traditional news media (e.g., CNN, BBC) rely on dedicated news reporters while social news aggregators (e.g., Digg, Reddit) rely on active voluntary participation of committed individuals. A key drawback of such news collection approaches is that they entrust a central authority (i.e., a news agency or web administrator) to analyze and verify disease outbreaks like COVID-19, which may induce delays in deriving the COVID-19 propagation [4]. In contrast, a decentralized social-sensing based news aggregation and subscription service can potentially accelerate the news collection as well as distribution of information during the global pandemic of COVID-19 [74]. A survey shows that 37% of Internet users promulgated news content through social media posts on Facebook and Twitter [74]. With the proliferation of smart devices and people's tendency to post about nearby cases of people showing symptoms of COVID-19 or feeling ill [75]–[77], information about probable COVID-19 cases can propagate very fast through the social media. However, as identified earlier, a key hurdle is to develop a system that can spontaneously locate, obtain, and store the data from the social media platforms. Furthermore, after the COVID-19 related information is assembled, a system needs to be developed that can convey the processed information to the mass public. A set of important research questions are: i) how to efficiently filter and organize information contributed by diversified and unreliable sources? ii) How to compile the gathered information to an acceptable degree that each subscriber feels complacent in reading and trusting? iii) How to present the information to less tech savvy individuals with limited knowledge on computers and smartphones? iv) How to sustain the news aggregation and circulation during an Internet downtime?

A possible approach of information collection is to develop a real-time social media data collection and storage engine, such as Apollo [78]. One other potentially effective technique for information aggregation is to develop a dedicated crowdsensing-based smartphone application which allows users to readily report about COVID-19 related observations [33]. Subsequently, a decentralized mesh network based news subscription service can be constructed from the collected data in the mobile app that is able to operate autonomously without a central authority. The service can be used to leverage the rich set of real-time observations of COVID-19 contained in the social data to explore the collective wisdom of common individuals without relying on dedicated news reporters. The entire service may be implemented within the aforementioned mobile app that can both

collect the information of the COVID-19 spread from the online users and also present the prepared news to others [33]. This process can virtually eliminate the existence of a central authority, hence reducing delays in information gathering and distribution in a CovidSens application.

D. Privacy Aware Location Discovery Based on Contextual Analysis

CovidSens applications are inherently location data driven and hence a potential domain of research in CovidSens can be to address the location data scarcity challenge from the social media data. Specifically, studies can focus on determining the location of the COVID-19 related report origination points in the absence of the geo-location metadata in the posts. We emphasize that during inferring the event report locations from the social media data, care must be given to respect individual privacy from the system perspective, which if done improperly may lead to serious privacy breaches. For example, while a user's location information may be deduced from the text data in social media, it may also be used to infer other sensitive information such as job, ethnicity, race, financial status. Leakage of these information may place users at risk and lead to loss of confidence in the developed system [61]. Therefore, one important area of research in CovidSense can focus on how to develop privacy-aware location inference tools based on the contextual analysis of social media data that protects the identity and privacy of the users.

Once the user privacy is ensured, a good amount of opportunity exists in designing techniques to leverage the contextual information that is embedded within the text content of a social media post (toponym resolution). Moreover, images contained with posts can also be useful in extrapolating an accurate estimate of the social media report's origination sites [79]. For example, an individual tweeting about COVID-19 symptoms claiming to be from a particular location can be given greater credibility if he or she posts with the image of the place. Another way to obtain the geo-location information of social media data can be to use image-based geocoding where subjects in the background of a posted image are cross-referenced with known landmarks or popular sites to find the location of the image [80].

People who post about disease symptoms in social media and "follow" other social media users with similar symptoms may be co-located [81]. Intuitively, if one user's location can be determined, the location of the related users may be discovered as well. However, individuals may also reside very far from one another. For instance, two friends showing COVID-19 related symptoms may be located in two different cities. Thus, additional features from the social media data may be analyzed to infer other evidence for being co-located. Rich privacy-aware location inference schemes can be developed that fuse friend-follower networks with the contextual information embedded within texts in tweets to determine the whereabouts of COVID-19 spread [62], [81]. An ensemble of solutions employing natural language processing (NLP) [82], deep neural networks (DNNs), and social network analysis can

be built to accurately infer the location information from the social media data [73], [79].

E. Integration of Social Sensing with Physical Sensing

As identified earlier, one key goal for developing effective CovidSens applications is to address the data reliability challenge stemming from the unreliable social media users. Beside uncertainty quantification, a strand of research to combat the data reliability challenge in CovidSens is to integrate social sensing with physical sensing paradigms (e.g., unmanned aerial vehicles (UAVs) and vehicular sensor networks (VSNs)) to verify the reports connected to COVID-19. Compared to UAVs and VSNs, social sensing has a broader outreach, but suffers from inconsistent reliability. On the other hand, UAVs and VSNs are fitted with arrays of sensors (e.g., temperature, humidity, and air quality sensors, cameras, microphones) [83] that allow them to sense COVID-19 related events with substantial fidelity [25]. However, they are limited in sensing scope and possess partial autonomy [24]. Leveraging the collective strengths of UAVs and VSNs with social sensing can potentially accelerate the discovery of COVID-19 related events. The reliable and high quality measurements provided by physical sensors naturally complement the uncertain estimation and broader sensing scope of social sensing. Driven by the social signals, the mobility and agility of UAVs and VSNs can allow them to be quickly sent to COVID-19 prone areas or hot zones to collect real-time evidence (e.g., people loitering on streets or gathering in larger groups) and ascertain whether the reported cases actually exists before sending out medical teams or law enforcement [83].

A few possible courses of work can focus on either integrating social sensing with unmanned aerial vehicles (UAVs), namely social drone [23], or with vehicular sensor networks (VSNs), namely social car [84] to sense the neighbourhood of COVID-19 affected areas for unwanted crowds, open pharmacies or emergency supply stores, and so on. A set of open research questions in these applications are: i) how to leverage the noisy social signals to quickly guide drones and cars to locations of interest? ii) How to accommodate various constraints imposed by the physical world (e.g., deadlines of urgent cases like dying patients, limited availability drones and their limited flight times)? iii) How to leverage the observations collected by the drones (e.g., unwanted crowds) to improve the social sensing process? Probable solutions that holistically solve the above challenges in the context of CovidSens systems are yet to be developed.

VI. CONCLUSION

In this paper, we introduce CovidSens, a new vision of reliable social sensing-based information distillation and risk alerting systems to monitor the COVID-19 spread and study the transmission dynamics of the contagious disease. We highlight a few key challenges in CovidSens applications including data collection, reliability, modality, presentation, and misinformation spread. By harnessing interdisciplinary techniques, CovidSens can combine the collective strengths

of social sensing with machine intelligence as well as human intelligence to perform real-time analyses on the obtained epidemiological data. CovidSense can yield more timely and accurate prediction of the COVID-19 spread which may subsequently be presented to end users through a collection of rich mobile apps and UAVs. We hope this paper will uphold CovidSens as an important avenue for guiding research to tackle the current COVID-19 pandemic around the world.

REFERENCES

[1] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher, "On truth discovery in social sensing: A maximum likelihood estimation approach," in *Proc. ACM/IEEE 11th Int Information Processing in Sensor Networks (IPSN) Conf*, Apr. 2012, pp. 233–244.

[2] "Coronavirus disease 2019 (covid-19) in the u.s." [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/cases-in-us.html

[3] "Coronavirus disease 2019 (covid-19) in the u.s." [Online]. Available: https://coronavirus.1point3acres.com/en

[4] D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji, and L. Kaplan, "The age of social sensing," *Computer*, vol. 52, no. 1, pp. 36–45, 2019.

[5] G. Haddow and K. Haddow, "Social media and the boston marathon bombings: A case study," *Physical Security & Emergency Management*, 2015.

[6] G. D. Haddow and K. S. Haddow, *Disaster communications in a changing media world*. Butterworth-Heinemann, 2013.

[7] R. Chunara, J. R. Andrews, and J. S. Brownstein, "Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera outbreak," *The American journal of tropical medicine and hygiene*, vol. 86, no. 1, pp. 39–45, 2012.

[8] "6.7 million people just mentioned the coronavirus on social media." [Online]. Available: https://t.co/be68zIJfYj?amp=1

[9] "Facebook, google discuss sharing smartphone data with government to fight coronavirus, but there are risks." [Online]. Available: https://www.cnbc.com/2020/03/19/facebook-google-could-share-smartphone-data-to-fight-coronavirus.html

[10] C. Smith, A. Mashhadi, and L. Capra, "Ubiquitous sensing for mapping poverty in developing countries," *Paper submitted to the Orange D4D Challenge*, 2013.

[11] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: universal patterns in human urban mobility," *PloS one*, vol. 7, no. 5, p. e37027, 2012.

[12] Y. Zhang, Y. Lu, D. Zhang, L. Shang, and D. Wang, "Risksens: A multi-view learning approach to identifying risky traffic locations in intelligent transportation systems using social and remote sensing," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1544–1553.

[13] M. T. Al Amin, T. Abdelzaher, D. Wang, and B. Szymanski, "Crowd-sensing with polarized sources," in *2014 IEEE International Conference on Distributed Computing in Sensor Systems*. IEEE, 2014, pp. 67–74.

[14] J. Marshall and D. Wang, "Mood-sensitive truth discovery for reliable recommendation systems in social sensing," in *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016, pp. 167–174.

[15] Y. Zhang, R. Zong, J. Han, H. Zheng, Q. Lou, D. Zhang, and D. Wang, "Transland: An adversarial transfer learning approach for migratable urban land usage classification using remote sensing," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 1567–1576.

[16] C. A. Boulton, H. Shotton, and H. T. Williams, "Using social media to detect and locate wildfires," in *Tenth International AAAI Conference on Web and Social Media*, 2016.

[17] D. Y. Zhang, D. Wang, H. Zheng, X. Mu, Q. Li, and Y. Zhang, "Large-scale point-of-interest category prediction using natural language processing models," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 1027–1032.

[18] D. Y. Zhang, Y. Zhang, Q. Li, T. Plummer, and D. Wang, "Crowdlearn: A crowd-ai hybrid system for deep learning-based damage assessment applications," in *Distributed Computing Systems (ICDCS), 2019 IEEE 39th International Conference on*. IEEE, 2019.

[19] D. Wang, T. Abdelzaher, and L. Kaplan, *Social sensing: building reliable systems on unreliable data*. Morgan Kaufmann, 2015.

[20] D. Zhang, D. Wang, N. Vance, Y. Zhang, and S. Mike, "On scalable and robust truth discovery in big data social media sensing applications," *IEEE Transactions on Big Data*, 2018.

[21] Z. Xu, H. Zhang, V. Sugumaran, K.-K. R. Choo, L. Mei, and Y. Zhu, "Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 44, 2016.

[22] D. Y. Zhang, D. Wang, and Y. Zhang, "Constraint-aware dynamic truth discovery in big data social media sensing," in *Big Data (Big Data), 2017 IEEE International Conference on*. IEEE, 2017, pp. 57–66.

[23] M. T. Rashid, D. Zhang, L. Shang, and D. Wang, "An integrated social media and drone sensing system for reliable disaster response," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020.

[24] M. T. Rashid, D. Y. Zhang, L. Shang, and D. Wang, "Sead: Towards a social-media-driven energy-aware drone sensing framework," in *2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2019, pp. 647–654.

[25] M. T. Rashid, D. Zhang, Z. Liu, H. Lin, and D. Wang, "Collabdrone: A collaborative spatiotemporal-aware drone sensing system driven by social sensing signals," in *2019 28th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2019, pp. 1–9.

[26] C. W. Schmidt, "Trending now: using social media to predict and track disease outbreaks," 2012.

[27] L. E. Charles-Smith, T. L. Reynolds, M. A. Cameron, M. Conway, E. H. Lau, J. M. Olsen, J. A. Pavlin, M. Shigematsu, L. C. Streichert, K. J. Suda *et al.*, "Using social media for actionable disease surveillance and outbreak management: a systematic literature review," *PloS one*, vol. 10, no. 10, 2015.

[28] N. Wilson, K. Mason, M. Tobias, M. Peacey, Q. Huang, and M. Baker, "Interpreting "google flu trends" data for pandemic h1n1 influenza: the new zealand experience," *Eurosurveillance*, vol. 14, no. 44, p. 19386, 2009.

[29] L. A. Kalogiros, K. Lagouvardos, S. Nikoletseas, N. Papadopoulos, and P. Tzamalis, "Allergymap: a hybrid mhealth mobile crowdsensing system for allergic diseases epidemiology: a multidisciplinary case study," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2018, pp. 597–602.

[30] M. Krieck, J. Dreesman, L. Otrusina, and K. Denecke, "A new age of public health: Identifying disease outbreaks by analyzing tweets," in *Proceedings of health web-science workshop, ACM Web Science Conference*, 2011, pp. 10–15.

[31] T. L. S. Chester, M. Taylor, J. Sandhu, S. Forsting, A. Ellis, R. Stirling, and E. Galanis, "Use of a web forum and an online questionnaire in the detection and investigation of an outbreak," *Online journal of public health informatics*, vol. 3, no. 1, 2011.

[32] J. S. Brownstein, C. C. Freifeld, B. Y. Reis, and K. D. Mandl, "Surveillance sans frontieres: Internet-based emerging infectious disease intelligence and the healthmap project," *PLoS medicine*, vol. 5, no. 7, 2008.

[33] C. C. Freifeld, K. D. Mandl, B. Y. Reis, and J. S. Brownstein, "Healthmap: global infectious disease monitoring through automated classification and visualization of internet media reports," *Journal of the American Medical Informatics Association*, vol. 15, no. 2, pp. 150–157, 2008.

[34] M. Toda, I. Njeru, D. Zurovac, S. O. Tipo, D. Kareko, M. Mwau, and K. Morita, "Effectiveness of a mobile short-message-service–based disease outbreak alert system in kenya," *Emerging infectious diseases*, vol. 22, no. 4, p. 711, 2016.

[35] V. L. Yu and L. C. Madoff, "Promed-mail: an early warning system for emerging diseases," *Clinical infectious diseases*, vol. 39, no. 2, pp. 227–232, 2004.

[36] M. Carter, "How twitter may have helped nigeria contain ebola," *BMJ: British Medical Journal (Online)*, vol. 349, 2014.

[37] K. Sun, J. Chen, and C. Viboud, "Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study," *The Lancet Digital Health*, 2020.

[38] A. Schiffmann, "Coronavirus dashboard." [Online]. Available: https://ncov2019.live/data

[39] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," *The Lancet Infectious Diseases*, 2020.

[40] "Covid-19 in us and canada." [Online]. Available: https://coronavirus.1point3acres.com/en

[41] "What we can learn from south korea and singapore's efforts to stop coronavirus (besides wearing face masks)." [Online]. Available: https://news.yahoo.com/m/9ff084c5-5c10-39ff-98c5-18cb5749a0fa/what-we-can-learn-from-south.html

[42] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, "Features, evaluation and treatment coronavirus (covid-19)," in *StatPearls [Internet]*. StatPearls Publishing, 2020.

[43] P. Haddawy, L. Frommberger, T. Kauppinen, G. De Felice, P. Charkratpahu, S. Saengpao, and P. Kanchanakitsakul, "Situation awareness in crowdsensing for disease surveillance in crisis situations," in *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*, 2015, pp. 1–5.

[44] S. Minaeian, J. Liu, and Y.-J. Son, "Vision-based target detection and localization via a team of cooperative uav and ugvs," *IEEE Transactions on systems, man, and cybernetics: systems*, vol. 46, no. 7, pp. 1005–1016, 2015.

[45] "Cops will start using drones fitted with night-vision cameras." [Online]. Available: https://www.the-sun.com/news/580338/cops-to-start-using-drones/

[46] M. A. Ruiz Estrada, "The uses of drones in case of massive epidemics contagious diseases relief humanitarian aid: Wuhan-covid-19 crisis," *Available at SSRN 3546547*, 2020.

[47] K. Nur'Aini, I. Najahaty, L. Hidayati, H. Murfi, and S. Nurrohmah, "Combination of singular value decomposition and k-means clustering methods for topic detection on twitter," in *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE, 2015, pp. 123–128.

[48] A. N. Jagannatha and H. Yu, "Structured prediction models for rnn based sequence labeling in clinical text," in *Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing*, vol. 2016. NIH Public Access, 2016, p. 856.

[49] F. M. Zanzotto, M. Pennacchiotti, and K. Tsioutsiouliklis, "Linguistic redundancy in twitter," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011, pp. 659–669.

[50] K. Makice, *Twitter API: Up and running: Learn how to build applications with the Twitter API.* " O'Reilly Media, Inc.", 2009.

[51] D. Wang, L. Kaplan, and T. F. Abdelzaher, "Maximum likelihood analysis of conflicting observations in social sensing," *ACM Transactions on Sensor Networks (ToSN)*, vol. 10, no. 2, p. 30, 2014.

[52] D. Wang, M. T. Amin, S. Li, T. Abdelzaher, L. Kaplan, S. Gu, C. Pan, H. Liu, C. C. Aggarwal, and R. Ganti, "Using humans as sensors: an estimation-theoretic perspective," in *Information Processing in Sensor Networks, IPSN-14 Proceedings of the 13th International Symposium on*. IEEE, 2014, pp. 35–46.

[53] D. Y. Zhang, R. Han, D. Wang, and C. Huang, "On robust truth discovery in sparse social media sensing," in *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2016, pp. 1076–1081.

[54] X. Yin, J. Han, and S. Y. Philip, "Truth discovery with multiple conflicting information providers on the web," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 6, pp. 796–808, 2008.

[55] D. Wang, "Big data and information distillation in social sensing," in *Big Data*. Auerbach Publications, 2016, pp. 137–158.

[56] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang, "Data cleaning: Overview and emerging challenges," in *Proceedings of the 2016 International Conference on Management of Data*, 2016, pp. 2201–2206.

[57] R. Birke, M. Bjoerkqvist, L. Y. Chen, E. Smirni, and T. Engbersen, "(big) data in a virtualized world: volume, velocity, and variety in cloud datacenters," in *12th {USENIX} Conference on File and Storage Technologies ({FAST} 14)*, 2014, pp. 177–189.

[58] Y. Zhang, N. Vance, D. Zhang, and D. Wang, "On opinion characterization in social sensing: A multi-view subspace learning approach," in *2018 14th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2018, pp. 155–162.

[59] N. Zhang, Y.-s. Chen, and J.-l. Wang, "Image parallel processing based on gpu," in *2010 2nd International Conference on Advanced Computer Control*, vol. 3. IEEE, 2010, pp. 367–370.

[60] "Allergy symptoms vs covid-19 symptoms." [Online]. Available: https://www.familyallergy.com/general/allergy-symptoms-vs-covid-19-symptoms/

[61] N. Vance, D. Y. Zhang, Y. Zhang, and D. Wang, "Privacy-aware edge computing in social sensing applications using ring signatures," in *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2018, pp. 755–762.

[62] C. Huang, D. Wang, and S. Zhu, "Where are you from: Home location profiling of crowd sensors from noisy and sparse crowdsourcing data," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[63] Y. Kim, J. Huang, and S. Emery, "Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection," *Journal of medical Internet research*, vol. 18, no. 2, p. e41, 2016.

[64] D. Fisman, E. Khoo, and A. Tuite, "Early epidemic dynamics of the west african 2014 ebola outbreak: estimates derived with a simple two-parameter model," *PLoS currents*, vol. 6, 2014.

[65] N. Wingfield, M. Isaac, and K. Benner, "Google and facebook take aim at fake news sites," *The New York Times*, vol. 11, p. 12, 2016.

[66] "Misinformation will undermine coronavirus responses." [Online]. Available: https://dailybrief.oxan.com/Analysis/DB250989/Misinformation-will-undermine-coronavirus-responses

[67] D. Wang, D. Zhang, Y. Zhang, M. T. Rashid, L. Shang, and N. Wei, "Social edge intelligence: Integrating human and artificial intelligence at the edge," in *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE, 2019, pp. 194–201.

[68] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal, "On credibility estimation tradeoffs in assured social sensing," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1026–1037, 2013.

[69] D. Wang and C. Huang, "Confidence-aware truth estimation in social sensing applications," in *International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2015, pp. 336–344.

[70] K. K. Kumar and G. Geethakumari, "Detecting misinformation in online social networks using cognitive psychology," *Human-centric Computing and Information Sciences*, vol. 4, no. 1, pp. 1–22, 2014.

[71] D. Y. Zhang, L. Shang, B. Geng, S. Lai, K. Li, H. Zhu, M. T. Amin, and D. Wang, "Fauxbuster: A content-free fauxtography detector using social media comments," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 891–900.

[72] D. Wang, T. Abdelzaher, L. Kaplan, R. Ganti, S. Hu, and H. Liu, "Exploitation of physical constraints for reliable social sensing," in *2013 IEEE 34th Real-Time Systems Symposium*. IEEE, 2013, pp. 212–223.

[73] D. Zhang, Y. Zhang, Q. Li, T. Plummer, and D. Wang, "Crowdlearn: A crowd-ai hybrid system for deep learning-based damage assessment applications," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 1221–1232.

[74] S. Hong, "Online news on twitter: Newspapers' social media adoption and their online readership," *Information Economics and Policy*, vol. 24, no. 1, pp. 69–74, 2012.

[75] R. Lamsal. (2020) Corona virus (covid-19) tweets dataset. [Online]. Available: http://dx.doi.org/10.21227/781w-ef42

[76] J. E. C. Saire and R. C. Navarro, "What is the people posting about symptoms related to coronavirus in bogota, colombia?" *arXiv preprint arXiv:2003.11159*, 2020.

[77] E. Chen, K. Lerman, and E. Ferrara, "Covid-19: The first public coronavirus twitter dataset," *arXiv preprint arXiv:2003.07372*, 2020.

[78] "Towards fact-finding for social (human-centric) sensing." [Online]. Available: https://apollo.cse.nd.edu/

[79] A. Gallagher, D. Joshi, J. Yu, and J. Luo, "Geo-location inference from image content and user tags," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2009, pp. 55–62.

[80] D. Lin, A. Kapoor, G. Hua, and S. Baker, "Joint people, event, and location recognition in personal photo collections using cross-domain context," in *European Conference on Computer Vision*. Springer, 2010, pp. 243–256.

[81] H. Gu, H. Hang, Q. Lv, and D. Grunwald, "Fusing text and friendships for location inference in online social networks," in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 1. IEEE, 2012, pp. 158–165.

[82] N. Dhavase and A. Bagade, "Location identification for crime & disaster events by geoparsing twitter," in *International Conference for Convergence for Technology-2014*. IEEE, 2014, pp. 1–3.

[83] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, "Help from the sky: Leveraging uavs for disaster management," *IEEE Pervasive Computing*, no. 1, pp. 24–32, 2017.

[84] M. T. Rashid, D. Zhang, and D. Wang, "Socialcar: A task allocation framework for social media driven vehicular network sensing systems," in *The 15th International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*. IEEE, 2019.