



CovidSens: a vision on reliable social sensing for COVID-19

Md Tahmid Rashid¹ · Dong Wang¹ 

Published online: 12 June 2020
© Springer Nature B.V. 2020

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 AI-driven sensing paradigm to extract real-time observations from online users. In this paper, we propose CovidSens, a vision of social sensing-based risk alert systems to spontaneously obtain and analyze social data to infer the state of the COVID-19 propagation. CovidSens can actively help to keep the general public informed about the COVID-19 spread and identify risk-prone areas by inferring future propagation patterns. The CovidSens concept is motivated by three observations: (1) people have been actively sharing their state of health and experience of the COVID-19 via online social media, (2) official warning channels and news agencies are relatively slower than people reporting their observations and experiences about COVID-19 on social media, and (3) online users are frequently equipped with substantially capable mobile devices that are able to perform non-trivial on-device computation for data processing and analytics. We envision an unprecedented opportunity to leverage the posts generated by the ordinary people to build a real-time sensing and analytic system for gathering and circulating vital information of the COVID-19 propagation. Specifically, the vision of CovidSens attempts to answer the questions: How to distill reliable information about the COVID-19 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? How to leverage the computational power on the edge devices (e.g., smartphones, IoT devices, UAVs) to construct fully integrated edge-based social sensing platforms for rapid detection of the COVID-19 spread? In this vision paper, we discuss the roles of CovidSens and identify the potential challenges in developing reliable social sensing-based risk alert systems. We envision that approaches originating from multiple disciplines (e.g., AI, 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.

Keywords Social sensing · COVID-19 · Coronavirus · Disease tracking · Real-time · Information distillation

✉ Dong Wang
dwang5@nd.edu

Extended author information available on the last page of the article

1 Introduction

In this era of big data and pervasive Internet connectivity, social sensing is emerging as a dynamic AI-driven sensing paradigm that utilizes observations by humans and devices coupled with powerful AI devices [e.g., dedicated AI system-on-a-chip (SOC)] to obtain information about the physical world (Ignatov et al. 2018; Wang et al. 2012a). In this vision paper, we present CovidSens, the notion of real-time risk analysis and alerting systems based on social sensing to obtain situational awareness and guide the intervention motives for the Coronavirus Disease 2019 (COVID-19). According to the most recent statistics, there are more than 1.5 million confirmed cases of COVID-19 and above of 89,660 deaths spread across 50 states in the US (Coronavirus disease 2019a, b). Most of the above cases happened within one week's time (i.e., between March 29, 2020 and April 04, 2020) and the current trend seems to be ever-increasing (Coronavirus disease 2019b). 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 (Vos and Buckner 2016). By contrast, information propagation across the social media and crowdsensing platforms is inherently faster than traditional news media (Wang et al. 2019a). For example, during the 2013 Boston Marathon Bombing, news about the first bomb explosion and the arrest of the suspect was posted on Twitter several minutes before news agencies made announcements (Haddow and Haddow 2015, 2013). After the onset of the Cholera outbreak in Haiti in 2010, the knowledge regarding the outbreak was first obtained from social media, which occurred weeks before officials confirmed the case of the outbreak (Chunara et al. 2012). Such cases exemplify the importance of social sensing during emergency scenarios such as now during the COVID-19 outbreak.

The CovidSens concept is thus motivated by three observations during this global crisis of COVID-19. First, 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, 6.7 million people talked about coronavirus on social **media.¹ Second, people report their observations on social media relatively faster than the official warning channels and news agencies that make formal announcements. As such, knowledge contribution and discovery through social sensing may offer more effective news transmission (Wang et al. 2019a). Third, online social media users, who report their observations of COVID-19, are frequently equipped with powerful mobile devices with rich processing capabilities (Ignatov et al. 2018). Such devices can execute complex AI models to distill information about the COVID-19 spread at the edge, potentially expediting the data analysis (Zhang et al. 2019a). Given these premises, we perceive an unprecedented opportunity to leverage the posts generated by the social media users to build a complete AI-driven analytics framework for rapidly 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 Fig. 1. These tweets express the experiences and observations of individuals

¹ 6.7 million people just mentioned the coronavirus on social media. <https://t.co/be68zJfYj?amp=1>.

Fig. 1 Tweets posted during the COVID-19 outbreak



about the COVID-19. If such tweets could be analyzed using state-of-the-art AI algorithms to identify regions affected by COVID-19 and determine the rate of the spread, it might potentially expedite the alleviation of the adverse effects of the virus. In addition, by parsing 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.² This could potentially help to divert people away from more crowded locations and hence reduce the spread of the disease.

While the CovidSens vision promises opportunities for a robust social sensing-based information distillation and alert service for the COVID-19 spread, several technical challenges exist in the way of building such a system to autonomously 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 based on adaptive AI architectures (Khan et al. 2019) that can extract credible information of the disease spread from the noisy, sparse, and unstructured social data contributed by unvetted human sources such as the tweets in Fig. 1. The third challenge exists in handling 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 how to distill information of the COVID-19 spread by customizing existing AI algorithms to run on the individually owned edge devices that are originally designed to run in a centralized fashion. The fifth challenge is to circulate the extracted information about the disease spread to the general public in a timely and

² Facebook, google discuss sharing smartphone data with government to fight coronavirus, but there are risks. <https://www.cncb.com/2020/03/19/facebook-google-could-share-smartphone-data-to-fight-coronavirus.html>.

efficient manner so that they can plan their actions accordingly. The sixth challenge lies in designing effective alert systems that consider the human aspect of the problem (i.e., handling people's reactions to alerts like fear, concern, or ignorance). The seventh 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 a dynamic and scalable AI-driven 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 inform about the COVID-19 spread using a combination 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 the 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 deep learning, machine learning, estimation theory, game theory, online social media analysis, distributed systems, 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 Sect. 2, we discuss a few state-of-the-art works in the direction of CovidSens. In Sect. 3, we explore potential real-world applications of CovidSens. We identify a few likely challenges in implementing a successful CovidSens system in Sect. 4. Afterward, in Sect. 5, 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 Sect. 6.

2 Related works

2.1 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 (Wang et al. 2019a). Examples of social sensing applications include predicting poverty in developing countries (Smith et al. 2013), studying human mobility in urban areas (Noulas et al. 2012), identifying traffic abnormalities (Zhang et al. 2020a; Wang et al. 2013a), monitoring the air quality (Zhang et al. 2019b), tracking social unrest (Al Amin et al. 2014) and disasters (Marshall and Wang 2016; Wang et al. 2013b), and detecting wild-fire (Boulton et al. 2016). A comprehensive survey of social sensing schemes is provided in Wang et al. (2015). Zhang et al. developed a scalable approach to obtain data veracity in social sensing (Zhang et al. 2018a). Xu et al. developed a framework for semantic and spatial analysis of urban emergency events using social media data (Xu et al. 2016). Zhang et al. presented a constraint-aware truth discovery model to detect dynamically evolving truth in social sensing (Zhang et al. 2017a). 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 (Rashid et al. 2020a). 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 a 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.

2.2 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 (Schmidt 2012; Charles-Smith et al. 2015). For example, Google launched a real-time influenza surveillance system, namely Google Flu Trends (Wilson et al. 2009), 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 (Kalogiros et al. 2018). Krieck *et al.* studied the possibility of analyzing Twitter data for infectious disease surveillance (Krieck et al. 2011). Chester et al. (2011) 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.

2.3 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 has necessitated the adoption of automated warning and alert systems (Schmidt 2012). Brownstein et al. contributed a few early works in this domain by developing: (i) a series of interactive websites, HealthMap and Flu Near You (Schmidt 2012; Brownstein et al. 2008), and (ii) a smartphone application called Outbreaks Near Me (Freifeld et al. 2008) 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 Toda et al. (2016). Yu et al. developed ProMED-mail, an early warning system for emerging diseases (Yu and Madoff 2004). Carter studied the possibility of a tweet-based information dispersal system to facilitate the containment of Ebola in Carter (2014). The above approaches are known to provide disease warnings 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.

2.4 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. (2020) proposed the first study that harnesses crowdsourced data from several social media sources to monitor the COVID-19 spread. Schiffmann³ developed an informative web portal that aggregates news from myriads of news sources to present the 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 (Dong et al. 2020). An online community of international students and professionals, called Ipoint3acres, developed a web-based real-time COVID-19 news aggregator to track the state of the spread in the US and Canada.⁴ A mobile app has been developed by the Singapore government to leverage crowdsourced information to locate community transmission of COVID-19.⁵ 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 (See Footnotes 3 and 4). 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.

2.5 AI-driven disease prediction

The growing demand for intelligent application domains like autonomous driving, robotics, computational medicine, computer vision, and natural language processing call for reliable AI-driven information distillation systems (Abiodun et al. 2018). In the recent past, several studies have used AI for diagnosis, identification, and monitoring of infectious diseases using data collected from various sources (e.g., past disease records, social media posts, wearable sensors) (Barrat et al. 2014; Kawtrakul et al. 2007; Torres et al. 2016). Babu et al. applied Grey Wolf optimization and recurrent neural networks (RNN) on patient symptom data for early disease detection and response (Babu et al. 2018). Du et al. proposed a convolutional neural networks (CNN)-based approach for measles risk identification by analyzing public perception of measles outbreak from Twitter data (Du et al. 2018). Torres et al. (2016) developed an artificial neural network (ANN)-based dengue tracking system based on prior infection data. Mahalakshmi et al. built a Zika virus outbreak prediction system from symptom data based on multilayer perception (MLP) neural networks (Mahalakshmi and Suseendran 2019). However, despite the usefulness of existing approaches, due to the lack of sufficiently sized datasets with high quality labels on COVID-19, a key concern in AI-driven COVID-19 detection is ending up with underfitted and biased AI models that could yield erroneous prediction (Naudé 2020). Moreover, while the above systems utilize efficient AI architectures for a prediction of specific diseases, they have not been tailored to handle the massive scale of the rapidly progressing COVID-19 spread that has heightened

³ Schiffmann A, Coronavirus dashboard. <https://ncov2019.live/data>.

⁴ Covid-19 in us and canada. <https://coronavirus.i3acres.com/en>.

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

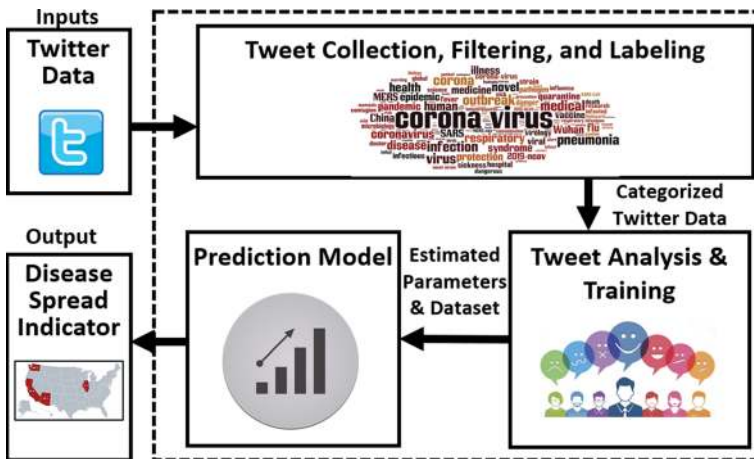


Fig. 2 Overview of an SDSI system

to a global pandemic. It is therefore a challenging task to develop scalable and adaptive real-time AI-based monitoring frameworks for COVID-19.

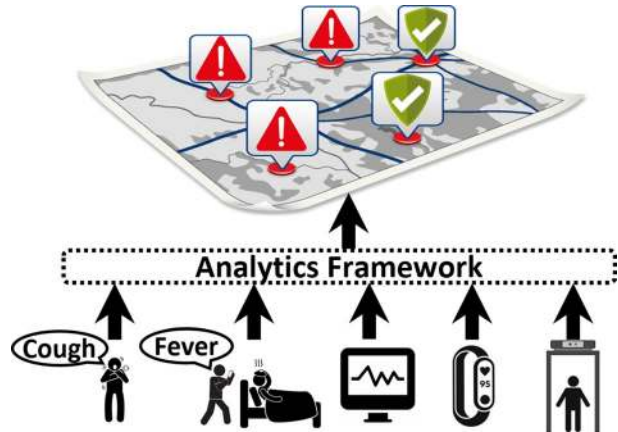
3 Real-world applications

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

3.1 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 (Sun et al. 2020). An example of an SDSI architecture is illustrated in Fig. 2. Initially, a real-time Twitter data crawler engine collects tweets indicating public opinions about the disease. The tweets are subsequently filtered and labeled 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) (Sun et al. 2020), (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 (Sun et al. 2020; Cascella et al. 2020). Afterward, 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 using adaptive AI algorithms such as Long Short Term Memory networks (LSTM) or Gated Recurrent Units (GRU) (Ma et al. 2016). 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

Fig. 3 Overview of a CDT system



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 deep learning techniques.

3.2 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) (Sun et al. 2020; Haddawy et al. 2015). 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 (Sun et al. 2020). In contrast to SDSI, CDT is relatively less pervasive and requires the 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 AI techniques on the obtained data to infer probable regions of infection and safe zones (Freifeld et al. 2008; Haddawy et al. 2015). To conserve bandwidth and expedite processing, the computational power of the smartphones can be harnessed to execute the AI algorithms at the edge. The third component is a smartphone application on the end-users' mobile phones to visually represent the analyzed geospatial distribution of the inferred regions (Freifeld et al. 2008). 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) (Zhang et al. 2018b). In most cases, the data collection, processing, and representation are carried out in the same smartphone application (Freifeld et al. 2008). Sun et al. proposed one of the earliest

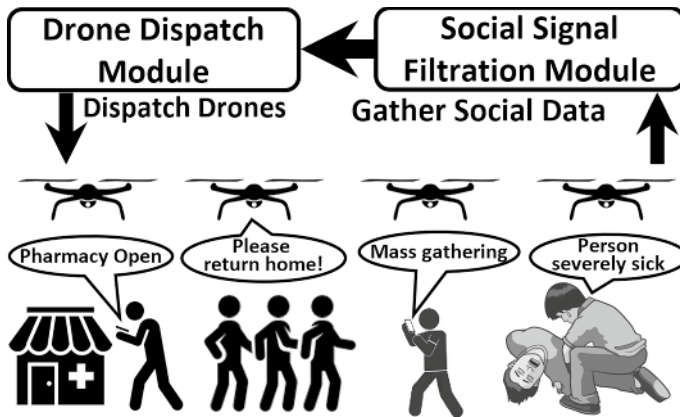


Fig. 4 Overview of an UHSA system

crowdsourcing based COVID-19 outbreak detection system (Sun et al. 2020). The Singapore and South Korea governments have launched mobile apps that utilize crowdsourced data to trace community transmission of the COVID-19 (See Footnote 5).

3.3 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 (Minaeian et al. 2015). 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 (Minaeian et al. 2015). 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. Afterward, the data is gathered in a backend server and processed using social sensing approaches based on statistical analysis, deep learning, 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 and image classification algorithms like Convolutional Neural Networks (CNNs), 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 and verify the availability of critical supplies using the UAVs (e.g., open pharmacy, grocery stores) based on the social media posts. Using the onboard speakers of the UAVs, 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 lockdown situation.⁶ During the COVID-19

⁶ Cops will start using drones fitted with night-vision cameras. <https://www.the-sun.com/news/580338/cops-to-start-using-drones/>.

crisis in China, UAVs have served multiple roles including post-epidemic aerial evaluation, alerting, and relief distribution to affected regions (Ruiz Estrada 2020).

4 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.

4.1 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 (Nur’Aini et al. 2015), adaptive sampling (Zhang et al. 2018c), and recurrent neural network (RNN) based textual labeling process (Jagannatha and Yu 2016). 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 (Zanzotto et al. 2011). On top of that, a good amount of social media data is observed to be transient and perishable (Zhang et al. 2019c). 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 (Makice 2009). 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.

4.2 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 (Wang et al. 2013c, 2014a, b, c). One important task while harnessing social media for CovidSens is to extract trustworthy information from unreliable human sources with unknown source reliability (Wang et al. 2012a). 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 (Wang et al. 2012a, 2014d). 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 (Zhang et al. 2018a, 2017a).

Yin et al. developed Truth Finder, a probabilistic algorithm using iterative weight updates to improve the quality of the data in social sensing (Yin et al. 2008). 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 (Zhang et al. 2017a). Existing methods are also not tailored towards disease outbreak detection, which may lead to a 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.

4.3 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 (Wang et al. 2015). 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 (Chu et al. 2016; Zhang et al. 2019d, 2020b; Shang et al. 2019a). 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 (Birke et al. 2014). 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 (Zhang et al. 2018d). This trend is advocated by the fact that image data processing involves heavy computation requirements (Zhang et al. 2010).

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.⁷ Fusing text with other data such as image and location data may potentially yield a 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 a greater computation power, care must be given to efficiently strike a 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 into 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)

⁷ Allergy symptoms vs covid-19 symptoms. <https://www.familyallergy.com/general/allergy-symptoms-vs-covid-19-symptoms/>.

How to speed up the analysis of multi-modal data for faster COVID-19 spread detection by distributing the computation across multiple devices?

4.4 AI-model scalability challenge

Due to the global scale of the COVID-19 outbreak, it is important to resort to adaptive AI-based methods that can effectively monitor the state of the spread from the social sensing data across any region of the world in real-time. This necessitates the scalable AI algorithms that can be readily deployed across the edge devices (e.g., smartphones, IoT devices, drones) in order to reduce latency and bandwidth consumption, and yield faster information extraction for the COVID-19 spread. Unfortunately, existing AI schemes such as DNNs, MLPs, and RNNs have been originally developed for powerful centralized hardware (e.g., GPU clusters) and are not tailored for resource-constrained smart devices residing at the edge of the network (Li et al. 2018; Zhang et al. 2019e, f). In particular, current AI algorithms are associated with model update processes that operate in a centralized fashion, which imposes a high network bandwidth requirement. In addition to that, mainstream AI models require extensive training to update the model parameters before being able to generate reliable predictions. Thus, even if the current AI algorithms could be improvised to run on the edge devices, due to their heavy computation requirements for the model training processes, they would drain the batteries of the portable edge devices faster (Vance et al. 2019; Zhang et al. 2018e, f). A few open questions in CovidSens originating from the AI-model scalability challenge are: (i) how to parallelize the AI model training process across the edge devices to speed up the model training and conserve network bandwidth? (ii) How to optimize the AI algorithms to run efficiently on the energy-constrained edge hardware? (iii) How to modularize the AI algorithms so that they can be seamlessly deployed across a large number of edge devices without a single point of failure?

4.5 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 (Vance et al. 2018). 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 (Zhang et al. 2018g, 2019g, h). 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 an 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 (Huang et al. 2017). However, such uni-dimensional approaches that rely 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.

4.6 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 (Schmidt 2012; Brownstein et al. 2008). 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 (Wang et al. 2019a). 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 (Wang et al. 2019b). 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 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.

4.7 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 (Kim et al. 2016). Such behavior 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 (Fisman et al. 2014). On the other extreme of the spectrum, we also acknowledge that a certain proportion of the population has 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.



Fig. 5 Tweets indicating fake news

4.8 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 (Yin et al. 2008). This has caused social media giants such as Facebook and Google to conduct worldwide campaigns to fight the propagation of fake news (Wingfield et al. 2016). 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.⁸ This phenomenon has been classified by WHO as an ‘infodemic’ (See Footnote 8). 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.

5 Road-map for future work

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

⁸ Misinformation will undermine coronavirus responses. <https://dailybrief.oxan.com/Analysis/DB250989/Misinformation-will-undermine-coronavirus-responses>.

5.1 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 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 (Wang et al. 2019b). 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 ages 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) (Wang and Huang 2015). 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 few 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) (Wang et al. 2013a, 2011a, b, 2012b). While these methods have been tested to operate optimally to provide the desired uncertainty quantification, it still remains 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.

5.2 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 originate from the behavior of individuals sharing what others share (Wang et al. 2014b; Kumar and Geethakumari 2014). 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 (Yin et al. 2008), (ii) analyzing textual evidence with associated images (Zhang et al. 2018h), and (iii) considering physical constraints and temporal dependencies of the evolving truth (Zhang et al. 2017a). 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 (Zhang et al. 2019i). 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 (Zhang et al. 2019i).

5.3 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 the active voluntary participation of committed individuals (Shang et al. 2019b). 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 (Wang et al. 2019a). 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 (Hong 2012). A survey shows that 37% of Internet users promulgated news content through social media posts on Facebook and Twitter (Hong 2012). With the proliferation of smart devices and people's tendency to post about being tested positive for COVID-19^{9,10,11}, as well as being tested positive on antibodies,¹² 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

⁹ You've tested positive for covid-19. who has a right to know? <https://time.com/5810231/covid19-sharing-positive-result/>.

¹⁰ What happens when you go coronavirus viral. <https://www.vox.com/recode/2020/4/21/21223701/coronavirus-viral-internet-twitter-facebook>.

¹¹ Coronavirus update: Covid-19 survivor chronicles journey on twitter, saying 'i'm so thankful to be alive'. <https://newyork.cbslocal.com/2020/04/10/coronavirus-survivor-story-david-lat/>.

¹² Belleville mayor has coronavirus antibodies, believes he had covid-19 months ago. <https://www.northjersey.com/story/news/esssex/belleville/2020/04/30/belleville-nj-mayor-tests-positive-coronavirus-antibodies/3057925001/>.

limited knowledge of computers and smartphones? (iv) How to sustain the news aggregation and circulation during an Internet downtime?

A possible approach to information collection is to develop a real-time social media data collection and storage engine, such as Apollo.¹³ One other potentially effective technique for information aggregation is to develop a dedicated crowdsensing-based smartphone application that allows users to readily report about COVID-19 related observations (Freifeld et al. 2008). 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 (Freifeld et al. 2008). This process can virtually eliminate the existence of a central authority, hence reducing delays in information gathering and distribution in a CovidSens application.

5.4 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 (Zhang et al. 2017b, 2019j). The leakage of this information may place users at risk and lead to a loss of confidence in the developed system (Vance et al. 2018). Therefore, one important area of research in CovidSens 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 (Gallagher et al. 2009). 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 (Lin et al. 2010).

People who post about disease symptoms in social media and "follow" other social media users with similar symptoms may be co-located (Gu et al. 2012). Intuitively, if one

¹³ Towards fact-finding for social (human-centric) sensing. <https://apollo.cse.nd.edu/>.

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 (Huang et al. 2017; Gu et al. 2012). An ensemble of solutions employing natural language processing (NLP) (Dhavase and Bagade 2014), deep neural networks (DNNs), and social network analysis can be built to accurately infer the location information from the social media data (Zhang et al. 2019i; Gallagher et al. 2009).

5.5 Edge intelligence with federated learning

One prospective domain for future research in CovidSens can be focused on addressing the scalability challenge of the AI models to effectively monitor the COVID-19 propagation from the social sensing data. In order to ensure that the most up-to-date information of the COVID-19 spread is available at any instant across any location, a large scale deployment of CovidSens is crucial. However, since traditional AI models are inherently built with a design philosophy that endorses centralized training (Zhang and Wang 2019), they may not be a viable approach for such a global scale implementation of CovidSens. Therefore, in order to reliably analyze the obtained data related to COVID-19 across a global extent, we envision expandable AI architectures that can be spontaneously deployed across a massive number of edge devices.

With the growth of powerful edge devices (e.g., smartphones, IoT devices) and the demand for distributed model training over a large number of computing nodes, federated learning (FL) is gaining traction as a distributed AI training paradigm (Konečný et al. 2016). In FL, a shared global AI model is trained from a collection of edge devices owned by end-users, while retaining the training data within the edge devices (Wang et al. 2019b). By not transmitting the private data to a central server, FL manages to preserve user privacy and therefore foster trust among the participating users. The principle of FL aligns appropriately with our vision of scalable social sensing systems by shifting AI from the cloud to edge devices. However, there are still several open challenges in FL that need to be addressed before establishing effective CovidSens systems. One recurring issue in FL is the inconsistent availability of the edge devices, otherwise known as *churn* (Vance et al. 2019). FL heavily relies on the participation of the edge devices for the training phase, which requires multiple iterations to converge to global optima. Edge devices are owned by rational individuals who might abruptly leave in the middle of an ongoing AI model training process (Wang et al. 2019b). Moreover, edge devices might periodically evict tasks for power savings, or have a higher priority task to supplant the model training task. This could potentially negate the learning process, yielding poor model parameter training (Vance et al. 2019). Another limitation of many existing FL schemes is that they rely on synchronous model update operations (Chen et al. 2019). At every iteration of the model training, the server aggregates the model weights after receiving updates from all the clients. Due to the heterogeneity of the edge devices and the instability of network connections, all the devices cannot be guaranteed to have the same update interval (Zhang et al. 2019e). Thus, the server is prone to substantial downtime while needing to wait for all local updates before aggregation. In a CovidSens application, where time is a crucial factor, such delays

are undesirable as they may slow the real-time prediction of the COVID-19 spread. Therefore, it is an open challenge to simultaneously handle the churn issue and develop asynchronous model training in FL for scalable CovidSens applications.

5.6 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) (Erdelj et al. 2017) that allow them to sense COVID-19 related events with substantial fidelity (Rashid et al. 2019a). However, they are limited in sensing scope and possess partial autonomy (Rashid et al. 2019b). 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 (Erdelj et al. 2017).

A few possible courses of work can focus on either integrating social sensing with UAVs, namely social drone (Rashid et al. 2020a), or with VSNs, namely social car (Rashid et al. 2019c) to sense the neighborhood of COVID-19 affected areas for unwanted crowds, open pharmacies or emergency supply stores, and so on. Social drone-based approaches can be further integrated with computational modeling (e.g., disease propagation models) to enhance the COVID-19 detection process (Rashid et al. 2020b). 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 and the limited availability of 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.

6 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, scalability, modality, presentation, and misinformation spread. By harnessing interdisciplinary techniques, CovidSens can combine the collective strengths of social sensing with AI as well as human

intelligence to perform real-time analyses on the obtained epidemiological data. Covid-Sense can yield a 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.

Acknowledgements This research is supported in part by the National Science Foundation under Grant Nos. CNS-1845639, CNS-1831669, Army Research Office under Grant No. W911NF-17-1-0409. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

- Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H (2018) State-of-the-art in artificial neural network applications: a survey. *Heliyon* 4(11):e00938
- Al Amin MT, Abdelzاهر T, Wang D, Szymanski B (2014) Crowd-sensing with polarized sources. In: 2014 IEEE international conference on distributed computing in sensor systems (IEEE, 2014), pp 67–74
- Babu SB, Suneetha A, Babu GC, Kumar YJN, Karuna G (2018) Medical disease prediction using grey wolf optimization and auto encoder based recurrent neural network. *Period Eng Nat Sci* 6(1):229
- Barrat A, Cattuto C, Tozzi AE, Vanhems P, Voirin N (2014) Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clin Microbiol Infect* 20(1):10
- Birke R, Bjoerkqvist M, Chen LY, Smirni E, Engbersen T (2014) (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
- Boulton CA, Shotton H, Williams HT (2016) Using social media to detect and locate wildfires. In: tenth international AAAI conference on web and social media
- Brownstein JS, Freifeld CC, Reis BY, Mandl KD (2008) Surveillance Sans Frontiers: internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med* 5(7):e151
- Carter M (2014) How Twitter may have helped Nigeria contain Ebola. *BMJ Br Med J* 349:g6946
- Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R (2020) Features, evaluation and treatment coronavirus (COVID-19). In: StatPearls. StatPearls Publishing, Treasure Island
- Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EH, Olsen JM, Pavlin JA, Shigematsu M, Streichert LC, Suda KJ et al (2015) Using social media for actionable disease surveillance and outbreak management: a systematic literature review. *PLoS ONE* 10(10):e0139701
- Chen Y, Sun X, Jin Y (2019) Communication-efficient federated deep learning with asynchronous model update and temporally weighted aggregation. [arXiv:1903.07424](https://arxiv.org/abs/1903.07424)
- Chester TLS, Taylor M, Sandhu J, Forsting S, Ellis A, Stirling R, Galanis E (2011) Use of a web forum and an online questionnaire in the detection and investigation of an outbreak. *Online J Public Health Inform* 3(1):ojphi.v3i1.3506
- Chu X, Ilyas IF, Krishnan S, Wang J (2016) Data cleaning: overview and emerging challenges. In: Proceedings of the 2016 international conference on management of data, pp. 2201–2206
- Chunara R, Andrews JR, Brownstein JS (2012) Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am J Trop Med Hyg* 86(1):39
- Coronavirus disease (2019a) (covid-19) in the U.S. <https://www.cdc.gov/coronavirus/2019-ncov/cases-in-us.html>
- Coronavirus disease (2019b) (covid-19) in the U.S. <https://coronavirus.1point3acres.com/en>
- Dhavase N, Bagade A (2014) Location identification for crime & disaster events by geoparsing Twitter. In: International conference for convergence for technology-2014 (IEEE, 2014), pp 1–3
- Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Du J, Tang L, Xiang Y, Zhi D, Xu J, Song HY, Tao C (2018) Public perception analysis of tweets during the 2015 measles outbreak: comparative study using convolutional neural network models. *J Med Internet Res* 20(7):e236

- Erdelj M, Natalizio E, Chowdhury KR, Akyildiz IF (2017) Help from the sky: leveraging UAVs for disaster management. *IEEE Pervasive Comput* 16(1):24
- Fisman D, Khoo E, Tuite A (2014) Early epidemic dynamics of the West African 2014 Ebola outbreak: estimates derived with a simple two-parameter model. *PLoS Curr* 8:6
- Freifeld CC, Mandl KD, Reis BY, Brownstein JS (2008) HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 15(2):150
- Gallagher A, Joshi D, Yu J, Luo J (2009) 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
- Gu H, Hang H, Lv Q, Grunwald D (2012) 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, (IEEE, 2012), vol. 1, pp 158–165
- Haddawy P, Frommberger L, Kauppinen T, De Felice G, Charkratpahu P, Saengpao S, Kanchanakitsakul P (2015) Situation awareness in crowdsensing for disease surveillance in crisis situations. In: Proceedings of the seventh international conference on information and communication technologies and development, pp 1–5
- Haddow GD, Haddow KS (2013) Disaster communications in a changing media world. Butterworth-Heinemann, Oxford
- Haddow G, Haddow K (2015) Social media and the Boston marathon bombings: a case study. *Physical Security & Emergency Management*
- Hong S (2012) Online news on Twitter: Newspapers' social media adoption and their online readership. *Inf Econ Policy* 24(1):69
- Huang C, Wang D, Zhu S (2017) 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
- Ignatov A, Timofte R, Chou W, Wang K, Wu M, Hartley T, Van Gool L (2018) Ai benchmark: running deep neural networks on android smartphones. In: Proceedings of the European conference on computer vision (ECCV) (2018)
- Jagannatha AN, Yu H (2016) 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. NIH Public Access, 2016, vol. 2016, p 856
- Kalogiros LA, Lagouvardos K, Nikolettseas S, Papadopoulos N, Tzamalís P (2018) 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
- Kawtrakul A, Yingsaeree C, Andres F (2007) A framework of NLP based information tracking and related knowledge organizing with topic maps. In: International conference on application of natural language to information systems. Springer, 2007, pp 272–283
- Khan A, Sohail A, Zahoor U, Qureshi AS (2019) A survey of the recent architectures of deep convolutional neural networks. [arXiv:1901.06032](https://arxiv.org/abs/1901.06032)
- Kim Y, Huang J, Emery S (2016) Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. *J Med Internet Res* 18(2):e41
- Konečný J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D (2016) Federated learning: Strategies for improving communication efficiency. [arXiv:1610.05492](https://arxiv.org/abs/1610.05492)
- Kriek M, Dreesman J, Otrusina L, Denecke K (2011) 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
- Kumar KK, Geethakumari G (2014) Detecting misinformation in online social networks using cognitive psychology. *Human-Centric Comput Inf Sci* 4(1):1
- Li H, Ota K, Dong M (2018) Learning IoT in edge: deep learning for the Internet of Things with edge computing. *IEEE Netw* 32(1):96
- Lin D, Kapoor A, Hua G, Baker S (2010) 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
- Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks. In: Proceedings of the 25th international joint conference on artificial intelligence (IJCAI 2016)


- Mahalakshmi B, Suseendran G (2019) Prediction of zika virus by multilayer perceptron neural network (MLPNN) using cloud. *Int J Recent Technol Eng (IJRTE)* 8:1–6
- Makice K (2009) *Twitter API: up and running: learn how to build applications with the Twitter API*. O'Reilly Media, Inc, Newton
- Marshall J, Wang D (2016) 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
- Minaeian S, Liu J, Son YJ (2015) Vision-based target detection and localization via a team of cooperative UAV and UGVs. *IEEE Trans Syst Man Cybern* 46(7):1005
- Naudé W (2020) Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. *Ai & Society* p 1
- Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) A tale of many cities: universal patterns in human urban mobility. *PLoS ONE* 7(5):e37027
- Nur'Aini K, Najahaty I, Hidayati L, Murfi H, Nurrohmah S (2015) 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
- Rashid MT, Zhang D, Liu Z, Lin H, Wang D (2019a) 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
- Rashid MT, Zhang DY, Shang L, Wang D (2019b) 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
- Rashid MT, Zhang D, Wang D (2019c) 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)*
- Rashid MT, Zhang D, Shang L, Wang D (2020a) An integrated social media and drone sensing system for Reliable Disaster Response. In: *IEEE INFOCOM 2020-IEEE conference on computer communications (IEEE 2020)*
- Rashid MT, Zhang Y, Zhang DY, Wang D (2020b) CompDrone: towards integrated computational model and social drone based wildfire monitoring. In: *16th international conference on distributed computing in sensor systems. (DCOSS20) (IEEE, 2020)*
- Ruiz Estrada MA (2020) The uses of drones in case of massive epidemics contagious diseases relief humanitarian aid: Wuhan-COVID-19 crisis
- Schmidt CW (2012) Trending now: using social media to predict and track disease outbreaks
- Shang L, Zhang DY, Wang M, Lai S, Wang D (2019a) Towards reliable online clickbait video detection: a content-agnostic approach. *Knowl-Based Syst* 182:104851
- Shang L, Zhang DY, Wang M, Wang D (2019b) VulnerCheck: a content-agnostic detector for online hatred-vulnerable videos. In: *2019 IEEE international conference on big data (big data) (IEEE, 2019)*, pp 573–582
- Smith C, Mashhadi A, Capra L (2013) Ubiquitous sensing for mapping poverty in developing countries. Paper submitted to the Orange D4D Challenge
- Sun K, Chen J, Viboud C (2020) Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study, *The Lancet Digital Health*
- Toda M, Njeru I, Zurovac D, Tipo SO, Kareko D, Mwau M, Morita K (2016) Effectiveness of a mobile short-message-service-based disease outbreak alert system in Kenya. *Emerg Infect Dis* 22(4):711
- Torres BY, Oliveira JHM, Tate AT, Rath P, Cumnock K, Schneider DS (2016) Tracking resilience to infections by mapping disease space. *PLoS Biol* 14(4):e1002436
- Vance N, Zhang DY, Zhang Y, Wang D (2018) 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
- Vance N, Rashid MT, Zhang D, Wang D (2019) Towards reliability in online high-churn edge computing: a deviceless pipelining approach. In: *2019 IEEE international conference on smart computing (SMARTCOMP) (IEEE, 2019)*, pp 301–308
- Vos SC, Buckner MM (2016) Social media messages in an emerging health crisis: tweeting bird flu. *J Health Commun* 21(3):301
- Wang D, Abdelzaher T, Kaplan L, Aggarwal CC (2011a) On quantifying the accuracy of maximum likelihood estimation of participant reliability in social sensing. In: *DMSN11: 8th international workshop on data management for sensor networks (2011)*
- Wang D, Abdelzaher T, Ahmadi H, Pasternack J, Roth D, Gupta M, Han J, Fatemeh O, Le H, Aggarwal CC (2011b) On bayesian interpretation of fact-finding in information networks. In: *14th international conference on information fusion (IEEE, 2011)*, pp 1–8

- Wang D, Kaplan L, Le H, Abdelzaher T (2012a) On truth discovery in social sensing: a maximum likelihood estimation approach. In: Proceedings of the ACM/IEEE 11th international conference on information processing in sensor networks (IPSN) (2012), pp 233–244. <https://doi.org/10.1109/IPSN.2012.6920960>
- Wang D, Kaplan L, Abdelzaher T, Aggarwal CC (2012b) On scalability and robustness limitations of real and asymptotic confidence bounds in social sensing. In: 2012 9th annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks (SECON) (IEEE, 2012), pp 506–514
- Wang D, Kaplan L, Abdelzaher T, Aggarwal CC (2013a) On credibility estimation tradeoffs in assured social sensing. *IEEE J Sel Areas Commun* 31(6):1026
- Wang D, Abdelzaher T, Kaplan L, Aggarwal CC (2013b) Recursive fact-finding: a streaming approach to truth estimation in crowdsourcing applications. In: 2013 IEEE 33rd international conference on distributed computing systems (IEEE, 2013), pp 530–539
- Wang D, Abdelzaher T, Kaplan L, Ganti R, Hu S, Liu H (2013c) Exploitation of physical constraints for reliable social sensing. In: 2013 IEEE 34th real-time systems symposium (IEEE, 2013), pp 212–223
- Wang D, Kaplan L, Abdelzaher TF (2014a) Maximum likelihood analysis of conflicting observations in social sensing. *ACM Trans Sensor Netw (ToSN)* 10(2):30
- Wang D, Al Amin MT, Abdelzaher T, Roth D, Voss CR, Kaplan LM, Tratz S, Laoudi J, Briesch D (2014b) Provenance-assisted classification in social networks. *IEEE J Select Topics Signal Process* 8(4):624
- Wang D, Abdelzaher T, Kaplan L (2014c) Surrogate mobile sensing. *IEEE Commun Mag* 52(8):36
- Wang D, Amin MT, Li S, Abdelzaher T, Kaplan L, Gu S, Pan C, Liu H, Aggarwal CC, Ganti R (2014d) Using humans as sensors: an estimation-theoretic perspective. In: Proceedings of the 13th international symposium on information processing in sensor networks, IPSN-14 (IEEE, 2014), pp 35–46
- Wang D, Huang C (2015) Confidence-aware truth estimation in social sensing applications. In: International conference on sensing, communication, and networking (SECON) (IEEE, 2015), pp 336–344
- Wang D, Abdelzaher T, Kaplan L (2015) Social sensing: building reliable systems on unreliable data. Morgan Kaufmann, Burlington
- Wang D, Szymanski BK, Abdelzaher T, Ji H, Kaplan L (2019a) The age of social sensing. *Computer* 52(1):36
- Wang D, Zhang D, Zhang Y, Rashid MT, Shang L, Wei N (2019b) 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
- Wilson N, Mason K, Tobias M, Peacey M, Huang Q, Baker M (2009) Interpreting “Google Flu Trends” data for pandemic H1N1 influenza: the New Zealand experience. *Eurosurveillance* 14(44):19386
- Wingfield N, Isaac M, Benner K (2016) Google and Facebook take aim at fake news sites. *N Y Times* 11:12
- Xu Z, Zhang H, Sugumaran V, Choo KKR, Mei L, Zhu Y (2016) Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media. *EURASIP J Wirel Commun Netw* 2016(1):44
- Yin X, Han J, Philip SY (2008) Truth discovery with multiple conflicting information providers on the web. *IEEE Trans Knowl Data Eng* 20(6):796
- Yu VL, Madoff LC (2004) ProMED-mail: an early warning system for emerging diseases. *Clin Infect Dis* 39(2):227
- Zanzotto FM, Pennacchiotti M, Tsioutsoulis K (2011) Linguistic redundancy in twitter. In: Proceedings of the conference on empirical methods in natural language processing (Association for Computational Linguistics, 2011), pp 659–669
- Zhang N, Chen Ys, Wang JI (2010) Image parallel processing based on GPU. In: 2010 2nd international conference on advanced computer control, vol. 3 (IEEE, 2010), pp 367–370
- Zhang DY, Wang D, Zhang Y (2017a) Constraint-aware dynamic truth discovery in big data social media sensing. In 2017 IEEE international conference on big data, IEEE, 2017, pp 57–66
- Zhang DY, Wang D, Zheng H, Mu X, Li Q, Zhang Y (2017b) 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
- Zhang D, Wang D, Vance N, Zhang Y, Mike S (2018a) On scalable and robust truth discovery in big data social media sensing applications. In: *IEEE transactions on big data*
- Zhang Y, Zhang D, Li Q, Wang D (2018b) Towards optimized online task allocation in cost-sensitive crowdsensing applications. In: 2018 IEEE 37th international performance computing and communications conference (IPCCC) (IEEE, 2018), pp 1–8
- Zhang Y, Zhang D, Vance N, Li Q, Wang D (2018c) A light-weight and quality-aware online adaptive sampling approach for streaming social sensing in cloud computing. In: 2018 IEEE 24th international conference on parallel and distributed systems (ICPADS) (IEEE, 2018), pp 1–8

- Zhang Y, Vance N, Zhang D, Wang D (2018d) 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
- Zhang D, Ma Y, Zhang Y, Lin S, Hu XS, Wang D (2018e) A real-time and non-cooperative task allocation framework for social sensing applications in edge computing systems. In: 2018 IEEE real-time and embedded technology and applications symposium (RTAS) (IEEE, 2018), pp 316–326
- Zhang D, Ma Y, Zheng C, Zhang Y, Hu XS, Wang D (2018f) Cooperative-competitive task allocation in edge computing for delay-sensitive social sensing. In: 2018 IEEE/ACM symposium on edge computing (SEC) (IEEE, 2018), pp 243–259
- Zhang Y, Lu Y, Zhang D, Shang L, Wang D (2018g) 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
- Zhang DY, Shang L, Geng B, Lai S, Li K, Zhu H, Amin MT, Wang D (2018h) 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
- Zhang DY, Wang D (2019) An integrated top-down and bottom-up task allocation approach in social sensing based edge computing systems. In: IEEE INFOCOM 2019-ieee conference on computer communications (IEEE, 2019), pp. 766–774
- Zhang D, Vance N, Wang D (2019a) When social sensing meets edge computing: vision and challenges. In: 2019 28th international conference on computer communication and networks (ICCCN), IEEE, 2019, pp 1–9
- Zhang Y, Zhang DY, Vance N, Wang D (2019b) An online reinforcement learning approach to quality-cost-aware task allocation for multi-attribute social sensing. *Pervasive Mobile Comput* 60:101086
- Zhang Y, Wang H, Zhang D, Wang D (2019c) Deeprisk: a deep transfer learning approach to migratable traffic risk estimation in intelligent transportation using social sensing. In: 2019 15th international conference on distributed computing in sensor systems (DCOSS) (IEEE, 2019), pp 123–130
- Zhang Y, Zong R, Han J, Zheng H, Lou Q, Zhang D, Wang D (2019d) 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
- Zhang D, Vance N, Zhang Y, Rashid MT, Wang D, Zhang D, Vance N, Zhang Y, Rashid MT, Wang D (2019e) In: 2019 IEEE Real-Time Systems Symposium (RTSS) (2019), pp 366–379
- Zhang D, Rashid T, Li X, Vance N, Wang D (2019f) Heteroedge: taming the heterogeneity of edge computing system in social sensing. In: Proceedings of the international conference on internet of things design and implementation (2019), pp 37–48
- Zhang Y, Wang H, Zhang D, Lu Y, Wang D (2019g) RiskCast: social sensing based traffic risk forecasting via inductive multi-view learning. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining (2019), pp 154–157
- Zhang Y, Dong X, Zhang D, Wang D (2019h) A syntax-based learning approach to geo-locating abnormal traffic events using social sensing. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining (2019), pp 663–670
- Zhang D, Zhang Y, Li Q, Plummer T, Wang D (2019i) 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
- Zhang D, Zhang Y, Li Q, Wang D (2019j) Sparse user check-in venue prediction by exploring latent decision contexts from location-based social networks. In: IEEE transactions on Big Data (2019)
- Zhang Y, Dong X, Shang L, Zhang D, Wang D (2020a) A multi-modal graph neural network approach to traffic risk forecasting in smart urban sensing. In: international conference on sensing, communication, and networking (SECON) (IEEE, 2020)
- Zhang Y, Zong R, Han J, Zhang D, Rashid T, Wang D (2020b) TransRes: a deep transfer learning approach to migratable image super-resolution in remote urban sensing. In: international conference on sensing, communication, and networking (SECON) (IEEE, 2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Md Tahmid Rashid¹ · Dong Wang¹ 

Md Tahmid Rashid
mrashid@nd.edu

¹ Department of Computer Science and Engineering, Notre Dame, IN 46556, USA