This is a pre-print version of the article. For published version please cite: Luis Fernandez-Luque, Muhammad Imran. Humanitarian Health Computing using Artificial Intelligence and Social Media: A Narrative Literature Review. International Journal of Medical Informatics. In press. Accepted 2018. DOI: https://doi.org/10.1016/j.ijmedinf.2018.01.015

Humanitarian Health Computing using Artificial Intelligence and Social Media: A Narrative Literature Review

Luis Fernandez-Luque, Muhammad Imran

Qatar Computing Research Institute, HBKU

Doha, Qatar

{*lluque, mimran*}*@hbku.edu.qa*

Abstract

Introduction: According to the World Health Organization (WHO), over 130 million people are in constant need of humanitarian assistance due to natural disasters, disease outbreaks, and conflicts, among other factors. These health crises can compromise the resilience of healthcare systems, which are essential for achieving the health objectives of the sustainable development goals (SDGs) of the United Nations (UN). During a humanitarian health crisis, rapid and informed decision making is required. This is often challenging due to information scarcity, limited resources, and strict time constraints. Moreover, the traditional approach to digital health development, which involves a substantial requirement analysis, a feasibility study, and deployment of technology, is ill-suited for many crisis contexts. The emergence of Web 2.0 technologies and social media platforms in the past decade, such as Twitter, has created a new paradigm of massive information and misinformation, in which new technologies need to be developed to aid rapid decision making during humanitarian health crises.

Objective: Humanitarian health crises increasingly require the analysis of massive amounts of information produced by different sources, such as social media content, and, hence, they are a prime case for the use of artificial intelligence (AI) techniques to help identify relevant information and make it actionable. To identify challenges and opportunities for using AI in humanitarian health crises, we reviewed the literature on the use of AI techniques to process social media.

Methodology: We performed a narrative literature review aimed at identifying examples of the use of AI in humanitarian health crises. Our search strategy was designed to get a broad overview of the different applications of AI in a humanitarian health crisis and their challenges. A total of 1,459 articles were screened, and 24 articles were included in the final analysis.
Results: Successful case studies of AI applications in a humanitarian health crisis have been reported, such as for outbreak detection. A commonly shared concern in the reviewed literature is the technical challenge of analyzing large amounts of data in real time. Data interoperability, which is essential to data sharing, is also a barrier with regard to the integration of online and traditional data sources.

Human and organizational aspects that might be key factors for the adoption of AI and social media remain understudied. There is also a publication bias toward high-income countries, as we identified few examples in low-income countries. Further, we did not identify any examples of certain types of major crisis, such armed conflicts, in which misinformation might be more common.
Conclusions: The feasibility of using AI to extract valuable information during a humanitarian health crisis is proven in many cases. There is a lack of research on how to integrate the use of AI into the work-flow and large-scale deployments of humanitarian aid during a health crisis.

Keywords: Health emergency, machine learning, global health, social media, internet, artificial intelligence, epidemiology

1. Introduction

According to the World Health Organization (WHO), over 130 million people are in need of humanitarian assistance, which can be related, among other factors, to natural disasters, disease outbreaks, and conflicts [1]. Managing a major health crisis requires addressing issues at various levels, from prevention and preparedness to response and recovery. Worldwide, and especially in low- to middle-income countries, a major health crisis can compromise the resilience of a country's healthcare systems, which are essential for achieving the health objectives of the sustainable development goals (SDGs).

Decision making during humanitarian health crises must be rapid and informed, and new technologies, especially those based on artificial intelligence (AI) techniques, are required; emerging data sources, such as social media, can facilitate this process [2]. Dealing with big data for a humanitarian health response is becoming increasingly necessary [3]; however, questions arise about how to apply AI techniques and utilize social media for humanitarian and global health issues.

This paper aims to provide an overview of the current state of the art in the use of AI and social media for humanitarian and global health. The purpose of this paper overlaps with various areas, including global health, crisis computing, humanitarian health, and digital health in low- to middle-income countries. The focus of this work is to study the role of AI in the application of digital health in those areas.

1.1. Background

Handling a health crisis is a crucial element of the SDGs. The third SDG, in particular, calls for strengthened capacity, research, and development to ensure health and well-being at any stage of life. Consequently, humanitarian health crises are a major risk that can threaten efforts toward ensuring health and well-being for all, at every stage of life. Medical informatics can help reduce the impact and risks inherent in such crises. Humanitarian health crises have inherent characteristics that must be considered: they (i) put stress on the resilience of health authorities (e.g., the situation can easily spiral out of control), (ii) might or might not be anticipated by relevant stakeholders (e.g., a hurricane season might be anticipated, but an outbreak of new disease might not), (iii) involve uncertainty due to the novelty of the crisis or a lack of previous knowledge and preparedness, and (iv) have an important impact on the health of a population and on the resilience of a country's healthcare system(s).

Humanitarian disasters create uncertainty but nevertheless demand rapid decision making, often with little or no information. Information scarcity is a major challenge when it comes to disaster response. However, with recent advances in information and communication technologies (ICT) and the emergence of social media sites, such as Twitter, researchers have tried to bridge this gap by incorporating information originating from multiple social media sources. AI techniques such as information retrieval, information classification and summarization, time-series analysis, and predictive analytics help process high volumes of data produced from heterogeneous data sources.

Building upon the above-described concepts of humanitarian health, AI and ICT techniques, and social media, we define the concept of *Humanitarian Health Computing (HHC)* as follows:

"The use of information and communication technologies (ICT) and artificial intelligence (AI) techniques to develop computational models for responses to humanitarian health crises caused by conflicts, wars, natural hazards, severe weather conditions, or disease outbreaks.

Overall, AI for HHC can be used for awareness, preparedness, prevention, and recovery. Possible outcomes of the use of AI for HHC include gaining situational awareness, learning about actionable cases, and improving communication throughout the lifespan of a health crisis. However, information is a key element for an effective application of AI techniques during a health crisis. Since decision making has to be done quickly and errors can be catastrophic, the availability of credible information is crucial. Big data can play a major role in health and humanitarian crises [4]. For example, mobile technologies have been used for monitoring food security during a humanitarian crisis [5]. As explained below, many applications of emerging social media data sources exist.

Among other uses, information from social media can aid the early detection of disease outbreaks. Studies have shown early reports on social media describing symptoms of a health problem [6, 7]. Once a potential outbreak is detected, keenly observing the situation as it unfolds is one of the core tasks of an effective healthcare response [8].

In crisis management, the importance of community and public education is emphasized. Decades after the HIV (Human immunodeficiency virus infection) outbreak, this epidemic is still a major public health concern; HHC could be a useful tool for engaging with the public in a scalable way [9, 10, 11].

Other emerging technologies that accord with our concept of HHC have been used to launch effective responses to humanitarian health crisis. For example, vulnerable populations can look for possible signs and symptoms of an ongoing outbreak, while affected individuals can ask treatment-related questions and visit nearby health emergency centers to get appropriate treatment [12].

The use of mobile and online data sources can easily lead to information overload, and manual processing of online information is not possible due to its high volume and velocity. During large outbreaks, the frequency of relevant social media messages could be up to several thousand per minute. Advanced data analytics approaches are being developed and used on platforms such as AIDR (Artificial Intelligence for Disaster Response) [13] or `http://healthmap.org/`. Related to this is the issue of detecting misinformation[14] and irrelevant content as most of the information available on social media during outbreaks could be irrelevant, and it often contains rumors rather than solid fact [15].

2. Methodology

Our literature review methodology is based on that of the scoping review (i.e., Tricco et al. [16]) and that designed by the International Medical Informatics Association (IMIA) Yearbook editors for surveying medical informatics sub-disciplines Lamy et al. [17]. We decided to base our review on those methods because their designs allowed us to get an overview of multiple disciplines. The use of a narrative review also allowed us to consider many types of article for the analysis.

2.1. Study selection

Figure 1 describes different steps performed for the selection of relevant articles.

Data sources and inclusion criteria: The inclusion criteria were set to include English language papers published from 1 January to 30 March 2017 in PubMed (292 results), the IEEE Xplore Digital Library (164 results), and the ACM Digital Library (1,051 results). There were 17 duplicates across all the databases. This selection criterion was followed to maximize the inclusion of both technical and health-related articles reporting on the use of AI in areas related to crisis computing for health and humanitarian health, such as feasibility studies, pilots, editorials, and reviews.

Generic query model and keyword selection: Our review spans several disciplines and also aims to provide an overview of the latest trends in the use of artificial intelligence in humanitarian health. For the design of the search queries, we followed a similar approach to the methodology created by Lame et. al. for the development of survey papers withing the IMIA (International Medical Informatics Association) Yearbook [17]. These surveys are designed to provide an overview of the current state of the art of sub-disciplines of medical informatics, taking into special consideration the multidisciplinary nature of the state of the art. Each query is a conjunction of four types of filters: (i) *topic filter*, (ii) *domain filter*, (iii) *situation filter*, and (iv) *publication filter*.

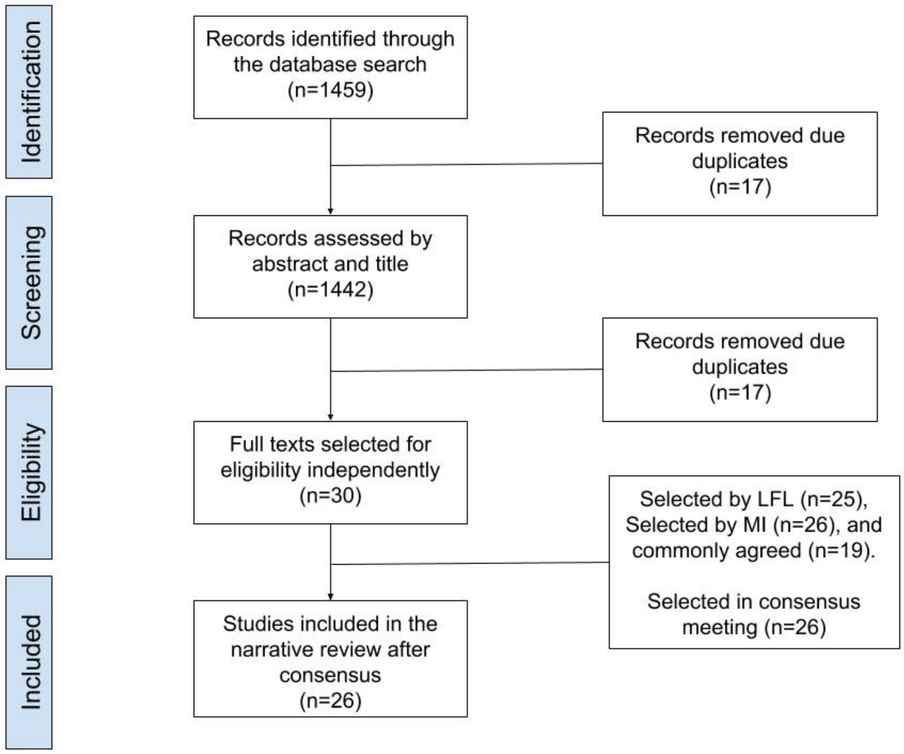

Figure 1: Flow diagram showing all the steps followed for the selection and inclusion of studies

The *domain* part of the query represents our domain, which is "health" during crises and emergencies. The *situation* part of the query brings crisis and humanitarian contexts such as wars, natural disasters etc. The *topic* part of the query includes technical keywords such as “machine learning”, “artificial intelligence”. Finally, we also included *publication filters* to ensure the retrieval of recent publications. As each scholarly database contains different types of articles and constraints, we adapted the generic structure of the query to a database-specific structure as follows (see table below).

For example, the PubMed query has the three filters as explained before: (i) topic filter about humanitarian and health crisis, (ii) domain filter, and (iii) publication filter. The queries of IEEE Xplore Digital Library and ACM Library were adapted to match the characteristics of corresponding databases.

Database	Query
IEEE Explorer	*(((health) AND (crisis OR humanitarian OR emergency OR war OR hurricane OR earthquake OR disaster OR volcan OR outbreak OR refugee OR tsunami OR cyclone OR Wildfire)* ***AND*** *(artificial intelligence OR machine learning OR deep learning OR computer vision)))*
PubMed	*("Health crisis"[Title/Abstract] OR "health emergency"[Title/Abstract] OR "Humanitarian"[Title/Abstract] OR "War"[Title/Abstract] OR "refugee"[Title/Abstract] OR "disaster"[Title/Abstract] OR Flood[Title/Abstract] OR Armed conflict [Title/Abstract] Internally displaced person[Title/Abstract] OR Hurricane[Title/Abstract] OR cyclone OR polio outbreak [Title/Abstract] tsunami[Title/Abstract] OR EARTHQUAKE[Title/Abstract] OR Volcan [Title/Abstract]OR Wildfire[Title/Abstract] OR Famine[Title/Abstract] OR Storm[Title/Abstract] OR Zika outbreak [Title/Abstract] OR Ebola outbreak[Title/Abstract] OR natural hazard[Title/Abstract] OR mers outbreak [Title/Abstract] OR tornado [Title/Abstract] OR extreme temperature [Title/Abstract] OR heat wave [Title/Abstract] OR cold wave [Title/Abstract] OR drought [Title/Abstract] OR pandemic [Title/Abstract] OR forest fire [Title/Abstract] OR nuclear hazard [Title/Abstract] OR nuclear explosion [Title/Abstract])* ***AND*** *("Machine learning"[Title/Abstract] OR "artificial intelligence"[Title/Abstract] OR "data mining"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "neuronal network"[Title/Abstract] OR "data analytics"[Title/Abstract] OR mobile[Title/Abstract] OR computing[Title/Abstract] OR internet[Title/Abstract] OR computer vision[Title/Abstract])* ***AND*** *((2017[DP] OR 2016[DP] OR 2015[DP] OR 2014[DP] OR 2013[DP]OR 2012[DP])*
ACM Library	*acmdlTitle:(+health crisis, emergency, humanitarian, war, refugee, disaster, conflict, flood, armed conflict, internally displaced person, hurricane, cyclone, polio outbreak, tsunami, earthquake, volcano, wildfire, famine, storm, zika outbreak, ebola outbreak, mers outbreak, natural hazard, tornado, blizzard, dust storm, extreme temperature, heat wave, cold wave, drought, forest fire, epidemic, pandemic, nuclear hazard, nuclear explosion)* ***AND*** *recordAbstract:(computing, artificial intelligence, ICT, machine learning, data mining, deep learning, neural network, data analytics, big data, informatics)* ***"filter"****: "publicationYear": "gte":2012, "lte":2017*

Study selection: Articles addressing the use of data mining in global and humanitarian health, which included social media as data sources, were included in the review. The authors reviewed a total of 1,459 articles independently using the application Rayyan Ouzzani et al. [18] based on each article's abstract and title. Of those, 17 duplicate articles were removed. The author LFL selected 25 articles, and MI selected 24, with agreement on 19 articles. Discrepancies were addressed in a consensus meeting which resulted in the inclusion of 26 articles.
Data extraction and analysis: Full-text articles were reviewed by both authors, and the main contributions of each paper were analyzed for this narrative review. For each study, we analyzed the aim, the target population, the health problem(s) being discussed, the technology involved, the country, the type of pilot study, the evaluation used, and the barriers and opportunities faced. Two papers were excluded from the full-text review: one de Quincey et al. [19] was excluded because it focused on hay fever and allergies, and a paper on disease risk mapping Raheja and S. Rajan [20] was excluded due to its lack of discussion of AI.

3. Results

As explained below, we mainly found publications dealing with the use of AI during major outbreaks, some of which were triggered due to natural disasters. Few articles reviewed the use of AI for applications in natural disasters (e.g., heat waves and pollution). Tables 1 and 2 show the classification of the reviewed literature into different health crisis phases and types.

Table 1: Studies classified into different phases of humanitarian health crises

Health crisis phases	Related studies
Prediction, early warning, and preparation	[21, 22, 23, 24, 25, 26]
Impact, damage assessment, and response	[27, 28, 29, 30, 31, 9]
Recovery and reconstruction	[32, 29, 30, 31, 33, 9]
Mitigation and prevention	[34, 35, 33, 36, 9, 24]

3.1. Surveys and systematic reviews

We decided to include reviews and surveys in our analysis to identify new areas of applications and potential barriers to the use of AI in global health and humanitarian aid crises.

Table 2: Studies classified into different health crisis types

Health crisis types	Related studies
Natural disasters	[25, 25]
Epidemics	[22, 37, 28] [27, 24] [38, 39, 21] [22, 23, 29] [32, 30, 31]
Pandemics	[40, 41]

A review published in 2012 provided an overview of the different steps involved in the text mining of online sources for digital epidemiology (mainly outbreak detection)[42]. Among the challenges highlighted in the article was the need for up-to-date ontologies that describe the content found in online sources and the need for the integration of offline data sources. This survey also provided an overview of freely available systems, such as HealthMap, BioCaster, and open source tools. Although the BioCaster project has been discontinued, its source code and ontology are freely available online at https://github.com/nhcollier/biocaster-ontology. In addition, the review of Saini and Kohli focused on the use of various machine learning techniques to analyze text from health social networks. A more recent review on the use of big data for health Fang et al. addressed technical issues in this area by emphasizing the difficulty of detecting potential outbreaks in real time.

L. Tsui et al. published a recent technical survey on the use of AI for tracking global pandemics [40]. The authors reviewed the necessary elements for modeling and forecasting pandemics using machine learning. They identified the challenge of integrating disparate data sources as one of the barriers to the development of this AI application, highlighting the need for more work on interoperability and policies to foster data sharing. The authors argued that more multidisciplinary work is needed that involves a wide range of stakeholders, including public health policymakers.

Al-garadi et al. provided a systematic review of the literature on pandemic surveillance using online social network data [41]. The authors stressed the rich information that social networks contain, which is useful to track pandemics. In the survey, a number of machine learning–based techniques for processing social media data are reported. The systematic review analyzed a total of 20 studies, but despite its focus on global pandemics, only two studies in low- to middle-income countries were identified (Brazil and China). Further, nearly all the studies focused on the use of Twitter, but this social network has low penetration in many low-income countries. A related review [38] reported the problem of false positives and negatives as one of the main challenges for the analysis of social media data in the detection of outbreaks.

Bates recently published an overview paper on digital epidemiology for predicting outbreaks [39]. In the article, the author explains the evolution of the discipline and describes some of the most well-known platforms, such as `http://healthmap.org/`. This overview also examines the barriers identified by leading researchers in the field. These include the privacy issues that underlie the use of online social media (e.g. privacy concerns of using shared personal information in social media). Another issue identified in the paper is the potential bias of digital epidemiology due to the lack of representation of some sectors of the population in social media. A review [45] focused on the social science aspects of digital epidemiology points out the disparity between medical terminology and the vocabulary used by laypersons to describe a symptom or condition. In this review, the authors explain transformations that occur at a social level, such as the active evolution of concepts, the incorporation of new informants to provide public health data, and the transformation of organizations, including the creation of health–data repositories.

3.2. Health outbreaks

3.2.1. General systems for outbreak detection

One of the most well-known platforms for outbreak detection is `http://healthmap.org/`, which is cited in many of the reviewed articles. Some articles also reported on other systems designed for outbreak detection. For example,Denecke et al. presents a system called M-Eco, which was primarily developed for epidemiologists, public health officials, or decision makers to monitor various information sources, such as social media, online news, TV, and radio, to detect emerging public health threats [21]. The system uses both supervised and unsupervised techniques to detect health issues and informs users about potential threats through a recommendation feature. In this case, the authors designed a classifier that facilitates the identification of content about potential threats.

Ji et al. proposed the monitoring of public concerns, emotions, and panic about health issues on social media as a public health surveillance tool. The authors presented the Epidemic Sentiment Monitoring System (ESMOS) to detect disease outbreaks over Twitter [22], employing sentiment classifiers to identify tweets with negative sentiment to generate a concern map and timeline chart.

In 2012, B. Neill published an article suggesting new technical approaches for the use of AI in outbreak detection [23]. The authors described how they used free text from emergency departments to create semantic scan statistics that could be used to identify topics that might be related to an outbreak. So, instead of looking for a specific disease, the authors developed a system based on topic discovery, which uses latent Dirichlet allocation.

3.2.2. Dengue vector control

One of the most crucial aspects of outbreak prevention is controlling vectors. For example, mosquito surveillance is used for the prevention of mosquito-borne diseases, such as malaria, yellow fever, and dengue. In Lee Chung-Hong et al.'s study, data about environmental risk factors for mosquito breeding are collected [34]. These include variables such as rainfall, humidity, and temperature, which were measured by sensors and collected and stored in online datasets. These datasets contained historically confirmed cases from health records in Taiwan. Using support vector machines (SVMs), the authors explored the feasibility of predicting dengue outbreaks based on online sensor data and explored the combination of SVM models with maps for spatiotemporal analysis.

3.2.3. Flu outbreak surveillance

Influenza is one of the most common infections in humans, and it is also a cause of major concern because it is highly contagious. In recent years, several strains of the influenza virus have caused global health emergencies, such as the H1N1 strain.

Chen et al. developed a topic model to predict the spread of influenza in South America using Twitter data, which included geographical cues to improve the accuracy of the models [27]. The analysis of web search logs is an additional strategy for detecting outbreaks. This strategy was explored by Araz et al., who found correlations between the search for flu-related terms in the area of Omaha (United States [US]) with an increasing number of visits to the local emergency department [28]. In another study, the researchers combined epidemiological data from the H1N1 flu outbreak with mobility data (e.g., air traffic information) to simulate the evolution of the outbreak on a global scale [37]. They reported on the use of influenza surveillance systems that harmonize data collection across public health agencies, such as www.epiwork.eu.

3.2.4. Ebola outbreak surveillance

The unprecedented Ebola outbreak in West Africa in 2014 and 2015 sparked a lot of research on the use of data-driven methods for responding to this health crisis. This research not only included the use of new technologies to better understand the outbreak, but also to simulate the potential impact of the outbreak if it were to reach other urban areas.

The Ebola outbreak also led to complex clinical encounters, where health professionals had to make decisions under stressful situations. Colubri et al. worked in a mobile system that integrated different data sources to predict an Ebola diagnosis for a given patient [29]. The system included data captured in mobile form, with clinical and laboratory data available from the EHR (Electronic Health Record). These predictive models were incorporated into the Ebola Computational Assignment of Risk Estimates (CARE) mobile application. The classifier built into the mobile phone was a single-layer artificial neural network due to the computing constraints of mobile devices.

In a study regarding the use of tweets to better understand the Ebola outbreak [32], the authors focused on the use of text mining to explore the public knowledge and attitudes of over 42 thousand tweets involving over 9 million users. They combined text mining with the geo-location of each tweet to explore those knowledge patterns across affected regions. The text mining methodology they used was outlined in a previous publication [30], which incorporated sentiment analysis from another study [31].

In [35], an simulation of Beijing, which consisted of a multilayer social network framework of about 19.6 million individuals with various real-life roles (e.g., infants, students, and workers) and 8 million buildings (e.g., workplaces, hospitals, and schools) [46], was used to reconstruct the spread of the Ebola epidemic following the propagation patterns observed in West Africa. Using machine learning models, which are mainly based on heuristics, the authors optimized the behaviors of individuals in the simulation by re-planning their daily activities (e.g., travel, sports, meals, and sleep). To predict epidemic situations for Ebola and influenza, two measures were used, i) infection probability and ii) contact frequency, and experiments were conducted under various durations (i.e., 100 days, 180 days, and 240 days). The researchers found that residential buildings were a main source of epidemic propagation, which means that families are possible carriers of infection.

3.2.5. The HIV pandemic: surveillance and education

HIV/AIDS continues to be a major global health issue, causing nearly thousands of deaths worldwide every year. Consequently, the prevention of HIV and support of those living with HIV continues to be part of the humanitarian effort for global health.

For the management of patients with HIV, a primary task is to identify the patients' demands for information. Thangarajan et al. collected and analyzed around 11 million geo-tagged tweets over a period of one year from the San Diego area in the US. Based on the HIV-related keywords generated by domain experts, the tweets were classified into five categories related to HIV using data mining techniques. Twitter-specific meta-information was used to generate a graph to identify relationships between users and their tweets. Ku et al. followed a similar approach, but instead of Twitter, they relied on web forums of people affected with HIV, using SVMs to interpret data from Yahoo answers [36].

The most effective measure to reduce the impact of HIV on global health is prevention, which requires health education. Imran and Castillo explored the use of the AI for Disaster Response (AIDR) platform [13] for the automatic classification of messages requesting HIV-related information; the platform is part of an SMS-type information center in Zambia run in cooperation with UNICEF. The goal of such classification was to improve the routing of HIV-related questions to the right health counselor [9, 24].

3.3. Natural disasters and humanitarian health

Despite the effects of natural disasters on human health, we found very few examples of the use of machine learning and AI in such cases. Caution is needed before reaching a conclusion; a possible explanation for this gap in the literature is that natural disasters might, in many cases, spark a health outbreak. For example, after the earthquake of Haiti in 2010, there was a cholera outbreak. Another issue related to the health of a population during a natural disaster is the potential destruction of roads, electricity infrastructure, and health facilities.

We identified one study that explored the use of AI to forecast the impact of environmental factors on health, such as the polluted air (i.e., smog) crisis in Beijing, China in 2013 [25]. In that study, the authors combined social media data from Twitter with ground sensor data and satellite images. The objective of the AI system was to predict smog-related health hazards using different machine learning algorithms (e.g., SVN, random forest, and artificial neural networks).

In Europe, the impact of heat waves on the population was a great concern, especially for the elderly. Keramitsoglou et al. et al. investigated the use of machine learning to forecast the risk of heat waves [26]. Heat waves are a growing concern due to the effects of climate change. Their model could predict heat wave hazards and their spatial distribution within large cities using hourly air temperature data taken from a thermal infrared satellite.

4. Discussion

Health organizations vary in terms of their information needs. Depending on their roles and responsibilities in addressing global health crises, an organization may look for different information than that of other organizations. Assessing, for example, the signs and symptoms of an unanticipated disease outbreak is a challenging task. Automatic approaches that were trained to work well for a past outbreak often fail to perform well in a new crisis. Many of the systems described in this literature review do not specifically address scalability and reuse of AI-based solutions across different crisis.Therefore, understanding these various needs for information is a crucial step for automatic AI-based systems to produce results that are suitable for a particular organization.

Social media is full of rumors and noisy content Westerman et al. [47]. However, the information processed by AI systems that will be used for decision-making purposes has to be credible, especially for healthcare. Determining the credibility of information originating from social media is a challenging task Castillo et al. [48]. Despite its importance, determining the credibility of information is an aspect that received the least attention in this literature review. Therefore, we emphasize the need for a credibility assessment component to be developed for future AI techniques, algorithms, and systems that are based on techniques and that address HHC issues. The need to tackle online misinformation has already been highlighted in policymaking discussions, such as the World Economic Forum Howell et al. [49], and can be also seen in current discussions on cybersecurity.

Information triage and triangulation processes address a number of issues related to refining data, such as completing missing information, addressing the interoperability of information systems that are communicating with each other, and information verification. For critical decision making, like in the domain of HHC, automatic processing systems should implement information triangulation strategies to improve the reliability of the results produced.

Furthermore, there is a lack of frameworks that facilitate data sharing Dye et al. [50]. In the context of a major crisis, many stakeholders start collecting and curating data, which, often, are not shared among different stakeholders. Many initiatives are emerging to solve these issues, such as `http://openmaps.org/`; however, formal data sharing frameworks and policies have to be established and enforced in the domain of HHC.

All the studies reporting on the use of social media data mentioned that one of the main challenges is the heterogeneity of social media content, as it contains much irrelevant information. In addition, there is a lack of public health data, which are essential to compare, combine, and complement online data sources.

More investigations of the human factors of HHC are needed, including user interfaces, data management policies, and capacity building. We found only one article addressing the social science aspects of global health French and Mykhalovskiy [45] despite its socio-ethical complexities Kickbusch [51]. Most previously published studies describe pilot projects, with very little information on what drives the use of AI in humanitarian and global health studies. For example, we did not find any previous research on the acceptance or usability of this kind of technology. Understanding these human factors would help us to determine which area of capacity building is necessary and how to implement an AI-system in real-life settings. Some manuscripts mentioned concerns about privacy Bates [39], but such concerns were not discussed in detail.

In our study, we focused only on published research in the literature. This could explain why we only found a few studies targeting low- to middle-income countries, despite the fact that many humanitarian and global health crises happen in such countries. This discrepancy is a well-known problem Haines et al. [52], which could not be easily overcome in our review.

For further development in this application area, the lack of a common framework for developing HHC techniques must be overcome. In our future work, we plan to use the findings, barriers, and shortcomings identified in this review to develop a framework that will guide the development and evaluation of new systems and technologies in the HHC domain.

5. Conclusions

Natural and man-made disasters pose serious challenges for communities and healthcare infrastructures in disaster areas. Among other factors, rapid access to information about victims and healthcare facilities can help to reduce suffering and rebuild communities. This review paper identified the usefulness of ICT-based technologies and AI techniques, when combined with online information sources like social media, to address humanitarian health issues. However, enormous challenges exist that must be overcome to fully utilize AI applications in the domain of HHC.

Online information sources (e.g., social networks) and data-driven AI approaches for health crises can help SDGs goals to be achieved. Nevertheless, guidelines and fully tested frameworks, as specified in the discussion section, are required before they can be used. The effective utilization of many technologies by stakeholders, which we have reviewed in this survey, has not yet been fully evaluated. This is mainly due to the lack of large-scale deployments and technology-driven pilot studies with a focus on human factor evaluation.

We identified a gap in the literature regarding the use of AI techniques for humanitarian health crises that do not involve an outbreak, and there is another gap in terms of AI and social media applications during humanitarian health crises caused by armed conflicts and natural disasters. Future research should address these gaps by building new technologies and systems that employ AI techniques and retrieve online information from social media.

6. References

[1] S. Akbarzada, T. K. Mackey, The syrian public health and humanitarian crisis: A displacementin global governance?, Global Public Health (2017) 1–17.

[2] T. K. Becker, B. Hansoti, S. Bartels, M. Bisanzo, G. A. Jacquet, K. Lunney, R. Marsh, M. Osei-Ampofo, I. Trehan, C. Lam, A. C. Levine, the Global Emergency Medicine Literature Review (GEMLR) Group, Global emergency medicine: A review of the literature from 2015, Academic Emergency Medicine 23 (2016) 1183–1191.

[3] F. Mulder, J. Ferguson, P. Groenewegen, K. Boersma, J. Wolbers, Questioning big data: Crowdsourcing crisis data towards an inclusive humanitarian response, Big Data & Society 3 (2016) 2053951716662054.

[4] C. Castillo, Big Crisis Data, Cambridge University Press, 2016.

[5] J.-M. Bauer, K. Akakpo, M. Enlund, S. Passeri, Tracking vulnerability in real time: Mobile text for food security surveys in eastern democratic republic of congo, Africa Policy Journal 9 (2013) 36.

[6] N. A. Christakis, J. H. Fowler, Social network sensors for early detection of contagious outbreaks, PloS one 5 (2010) e12948.

[7] R. Chunara, J. R. Andrews, 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 86 (2012) 39–45.

[8] S. Vieweg, A. L. Hughes, K. Starbird, L. Palen, Microblogging during two natural hazards events: what twitter may contribute to situational awareness (2010) 1079–1088.

[9] M. Imran, C. Castillo, Volunteer-powered automatic classification of social media messages for public health in aidr, Proceedings of the 23rd International Conference on World Wide Web 0 (2014) 671–672.

[10] R. M. Merchant, S. Elmer, N. Lurie, Integrating social media into emergency-preparedness efforts, New England Journal of Medicine 365 (2011) 289–291.

[11] K. Denecke, S. Atique, Social media and health crisis communication during epidemics, Participatory Health through Social Media (2016) 42.

[12] C. Castillo, F. Diaz, H. Purohit, Leveraging social media and web of data for crisis response coordination (2014).

[13] M. Imran, C. Castillo, J. Lucas, P. Meier, S. Vieweg, Aidr: Artificial intelligence for disaster response, in: Proceedings of the 23rd International Conference on World Wide Web, ACM, pp. 159–162.

[14] P. J. Auter, A. Douai, H. Makady, C. West, Circulating health rumors in the arab world: A 12-month content analysis of news stories and reader commentary about middle east respiratory syndrome from two middle eastern news outlets, International Communication Gazette 78 (2016) 411–431.

[15] M. Imran, C. Castillo, F. Diaz, S. Vieweg, Processing social media messages in mass emergency: A survey, ACM Computing Surveys (CSUR) 47 (2015) 67.

[16] A. C. Tricco, J. Antony, W. Zarin, L. Strifler, M. Ghassemi, J. Ivory, L. Perrier, B. Hutton, D. Moher, S. E. Straus, A scoping review of rapid review methods, BMC medicine 13 (2015) 224.

[17] J. Lamy, B. Séroussi, N. Griffon, G. Kerdelhué, M. Jaulent, J. Bouaud, et al., Toward a formalization of the process to select imia yearbook best papers, Methods Inf Med 54 (2015) 135–44.

[18] M. Ouzzani, H. Hammady, Z. Fedorowicz, A. Elmagarmid, Rayyana web and mobile app for systematic reviews, Systematic Reviews 5 (2016) 210.

[19] E. de Quincey, T. Kyriacou, T. Pantin, # hayfever; a longitudinal study into hay fever related tweets in the uk, in: Proceedings of the 6th International Conference on Digital Health Conference, ACM, pp. 85–89.

[20] V. Raheja, K. S. Rajan, Risk analysis based on spatio-temporal characterization: a case study of disease risk mapping, Proceedings of the First ACM SIGSPATIAL International Workshop on Use of GIS in Public Health 0 (2012) 48–56.

[21] K. Denecke, P. Dolog, P. Smrz, Making use of social media data in public health, in: Proceedings of the 21st International Conference on World Wide Web, ACM, pp. 243–246.

[22] X. Ji, S. A. Chun, J. Geller, Monitoring public health concerns using twitter sentiment classifications, in: Healthcare Informatics (ICHI), 2013 IEEE International Conference on, IEEE, pp. 335–344.

[23] D. B. Neill, New directions in artificial intelligence for public health surveillance, IEEE Intelligent Systems 27 (2012) 56–59.

[24] M. Imran, P. Meier, C. Castillo, A. Lesa, M. Garcia Herranz, Enabling digital health by automatic classification of short messages, Proceedings of the 6th International Conference on Digital Health Conference 0 (2016) 61–65.

[25] J. Chen, H. Chen, D. Hu, J. Z. Pan, Y. Zhou, Smog disaster forecasting using social web data and physical sensor data, 2015 IEEE International Conference on Big Data (Big Data) 0 (2015) 991–998.

[26] I. Keramitsoglou, C. T. Kiranoudis, B. Maiheu, K. De Ridder, I. A. Daglis, P. Manunta, M. Paganini, Heat wave hazard classification and risk assessment using artificial intelligence fuzzy logic., Environmental monitoring and assessment 185 (2013) 8239–58.

[27] L. Chen, K. S. M. T. Hossain, P. Butler, N. Ramakrishnan, B. A. Prakash, Flu gone viral: Syndromic surveillance of flu on twitter using temporal topic models, 2014 IEEE International Conference on Data Mining 0 (2014) 755–760.

[28] O. M. Araz, D. Bentley, R. L. Muelleman, Using google flu trends data in forecasting influenza-like-illness related ed visits in omaha, nebraska., The American journal of emergency medicine 32 (2014) 1016–23.

[29] A. Colubri, P. Sabeti, T. Silver, T. Fradet, K. Retzepi, B. Fry, Transforming clinical data into actionable prognosis models: Machine-learning framework and field-deployable app to predict outcome of ebola patients., PLoS Neglected Tropical Diseases [electronic resource] 10 (2016) e0004549.

[30] S. Yoon, N. Elhadad, S. Bakken, A practical approach for content mining of tweets, American journal of preventive medicine 45 (2013) 122–129.

[31] A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, Processing (2009) 1–6.

[32] S. Yoon, M. Odlum, What can we learn about the ebola outbreak from tweets?, American journal of infection control 43 (2015) 563–71.

[33] N. Thangarajan, N. Green, A. Gupta, S. Little, N. Weibel, Analyzing social media to characterize local hiv at-risk populations, in: Proceedings of the conference on Wireless Health, ACM, p. 11.

[34] C. H. Lee, H. C. Yang, S. J. Lin, Incorporating big data and social sensors in a novel early warning system of dengue outbreaks, 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) 0 (2015) 1428–1433.

[35] P. Zhang, B. Chen, L. Ma, Z. Li, Z. Song, W. Duan, X. Qiu, The large scale machine learning in an artificial society: prediction of the ebola outbreak in beijing, Computational intelligence and neuroscience 2015 (2015) 6.

[36] Y. Ku, C. Chiu, Y. Zhang, L. Fan, H. Chen, Global disease surveillance using social media: Hiv/aids content intervention in web forums, 2010 IEEE International Conference on Intelligence and Security Informatics 0 (2010) 170–170.

[37] M. Tizzoni, P. Bajardi, C. Poletto, J. J. Ramasco, D. Balcan, B. Gonalves, N. Perra, V. Colizza, A. Vespignani, Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm., BMC medicine 10 (2012) 165.

[38] I. Young, M. T. Pham, T. M. Bernardo, A. Rajic, K. Robiadek, J. A. Funk, Scoping review on search queries and social media for disease surveillance: a chronology of innovation., Journal Of Medical Internet Research 15 (2013) e147.

[39] M. Bates, Tracking disease: Digital epidemiology offers new promise in predicting outbreaks, IEEE Pulse 8 (2017) 18–22.

[40] K. L. Tsui, Z. S. Y. Wong, D. Goldsman, M. Edesess, Tracking infectious disease spread for global pandemic containment, IEEE Intelligent Systems 28 (2013) 60–64.

[41] M. A. Al-garadi, M. S. Khan, K. D. Varathan, G. Mujtaba, A. M. Al-Kabsi, Using online social networks to track a pandemic: A systematic review, Journal of biomedical informatics 62 (2016) 1–11.

[42] N. Collier, Uncovering text mining: a survey of current work on web-based epidemic intelligence., Global public health 7 (2012) 731–49.

[43] S. Saini, S. Kohli, Machine learning techniques for effective text analysis of social network e-health data, 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) 0 (2016) 3783–3788.

[44] R. Fang, S. Pouyanfar, Y. Yang, S.-C. Chen, S. S. Iyengar, Computational health informatics in the big data age: A survey, ACM Comput. Surv. 49 (2016) 1–36.

[45] M. French, E. Mykhalovskiy, Public health intelligence and the detection of potential pandemics., Sociology of health illness 35 (2013) 174–87.

[46] Y. Ge, L. Liu, X. Qiu, H. Song, Y. Wang, K. Huang, A framework of multilayer social networks for communication behavior with agent-based modeling, Simulation 89 (2013) 810–828.

[47] D. Westerman, P. R. Spence, B. Van Der Heide, Social media as information source: Recency of updates and credibility of information, Journal of Computer-Mediated Communication 19 (2014) 171–183.

[48] C. Castillo, M. Mendoza, B. Poblete, Predicting information credibility in time-sensitive social media, Internet Research 23 (2013) 560–588.

[49] L. Howell, et al., Digital wildfires in a hyperconnected world, World Economic Forum Report 3 (2013) 15–94.

[50] C. Dye, K. Bartolomeos, V. Moorthy, M. P. Kieny, Data sharing in public health emergencies: A call to researchers, Bulletin of the World Health Organization 94 (2016) 158.

[51] I. Kickbusch, Global health governance challenges 2016–are we ready?, International journal of health policy and management 5 (2016) 349.

[52] A. Haines, S. Kuruvilla, M. Borchert, Bridging the implementation gap between knowledge and action for health, Bulletin of the World Health Organization 82 (2004) 724–731.